

Searching for Quasi-periodic Oscillations in Astrophysical Transients Using Gaussian Processes

Moritz Hübner^{1,2} , Daniela Huppenkothen³ , Paul D. Lasky^{2,4} , Andrew R. Inglis^{5,6} , Christopher Ick⁷, and David W. Hogg⁸ 

¹ School of Physics, University of Melbourne, Parkville, VIC 3010, Australia; moritz.thomas.huebner@gmail.com

² OzGrav: The ARC Centre of Excellence for Gravitational Wave Discovery, Clayton, VIC 3800, Australia

³ SRON Netherlands Institute for Space Research, Niels Bohrlaan 4, 2333 CA Leiden, The Netherlands; d.huppenkothen@sron.nl

⁴ School of Physics and Astronomy, Monash University, Clayton, VIC 3800, Australia

⁵ Physics Department, The Catholic University of America, Washington, DC 20064, USA

⁶ Solar Physics Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA

⁷ Music and Audio Research Laboratory, New York University, 370 Jay St., Brooklyn, NY 11201, USA

⁸ Center for Cosmology and Particle Physics, Department of Physics, New York University, 726 Broadway, New York, NY 10003, USA

Received 2022 May 2; revised 2022 June 10; accepted 2022 June 14; published 2022 August 25

Abstract

Analyses of quasi-periodic oscillations (QPOs) are important to understanding the dynamic behavior in many astrophysical objects during transient events like gamma-ray bursts, solar flares, magnetar flares, and fast radio bursts. Astrophysicists often search for QPOs with frequency-domain methods such as (Lomb–Scargle) periodograms, which generally assume power-law models plus some excess around the QPO frequency. Time-series data can alternatively be investigated directly in the time domain using Gaussian process (GP) regression. While GP regression is computationally expensive in the general case, the properties of astrophysical data and models allow fast likelihood strategies. Heteroscedasticity and nonstationarity in data have been shown to cause bias in periodogram-based analyses. GPs can take account of these properties. Using GPs, we model QPOs as a stochastic process on top of a deterministic flare shape. Using Bayesian inference, we demonstrate how to infer GP hyperparameters and assign them physical meaning, such as the QPO frequency. We also perform model selection between QPOs and alternative models such as red noise and show that this can be used to reliably find QPOs. This method is easily applicable to a variety of different astrophysical data sets. We demonstrate the use of this method on a range of short transients: a gamma-ray burst, a magnetar flare, a magnetar giant flare, and simulated solar flare data.

Unified Astronomy Thesaurus concepts: [Astrostatistics \(1882\)](#); [Bayesian statistics \(1900\)](#); [Astrostatistics techniques \(1886\)](#); [Model selection \(1912\)](#); [X-ray transient sources \(1852\)](#); [Gamma-ray transient sources \(1853\)](#); [Magnetars \(992\)](#); [Solar flares \(1496\)](#); [Gamma-ray bursts \(629\)](#); [Time series analysis \(1916\)](#)

1. Introduction

In the analysis of astrophysical time series, quasi-periodic oscillations (QPOs) are peaks of excess energy in the power spectrum of the time series, corresponding to semicoherent oscillations where the amplitude, the oscillation frequencies, or both may jitter. QPOs have been observed in many different astrophysical sources, such as giant magnetar flares (Israel et al. 2005; Strohmayer & Watts 2005, 2006; Watts & Strohmayer 2006; Huppenkothen et al. 2014b; Miller et al. 2019; Castro-Tirado et al. 2021), X-ray binaries (see Ingram & Motta 2019 for a review), and solar flares (see Nakariakov & Melnikov 2009; Zimovets et al. 2021 for reviews).

They provide an exciting avenue to probe the properties of astrophysical objects. For example, QPOs in giant magnetar flares are likely associated with magnetic coupling of the crust to the core (Levin 2006, 2007; Colaiuda et al. 2009; Gabler et al. 2011; Levin & van Hoven 2011; Gabler et al. 2013). Identifying torsional modes may eventually shed light on neutron star properties (Samuelsson & Andersson 2007; Sotani et al. 2007; Colaiuda & Kokkotas 2011; Sotani et al. 2016, 2017). QPOs may

also eventually uncover the nature and properties of the central engine of gamma-ray bursts (GRBs; Ziaeepour & Gardner 2011). In solar flares, it is not yet firmly established whether QPOs originate from magnetohydrodynamic oscillations in coronal structures or due to magnetic reconnection (Zimovets et al. 2021), and understanding these processes will also shed light on the origin of stellar flares.

Most common analyses of time series that look for QPOs use frequency-domain-based methods such as periodograms or power spectra (Israel et al. 2005; Strohmayer & Watts 2005, 2006; Watts & Strohmayer 2006; Cenko et al. 2010; De Luca et al. 2010; Huppenkothen et al. 2012, 2014a, 2014b, 2015; Inglis et al. 2015, 2016; Auchère et al. 2016; Huppenkothen et al. 2017; Broomhall et al. 2019; Hayes et al. 2019, 2020; Miller et al. 2019; Andersen et al. 2021; Castro-Tirado et al. 2021; Tarnopolski & Marchenko 2021; Pastor-Marazuela et al. 2022). One performs analysis under the assumption that bins in a periodogram are χ^2 -distributed; the corresponding likelihood for an infinitely long time series with homoscedastic stationary Gaussian noise is called the Whittle likelihood (Whittle 1951). The Whittle likelihood has been used extensively in astronomy to infer parameters of models of stationary stochastic processes, for example, in active galactic nuclei (AGNs) and X-ray binaries (Barret & Vaughan 2012). The assumption of stationarity in particular is an issue for rapidly varying, nonstationary time series such as those observed from



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

fast transients like GRBs, magnetar bursts, and, recently, fast radio bursts (FRBs), where the total duration of the burst is perhaps only a few times longer than a candidate QPO period. Previous work has shown that for fast transients the statistical distributions at low frequencies deviate significantly from those assumed for a stationary process (Huppenkothen et al. 2012), and that assuming a Whittle likelihood can in some circumstances severely overstate the actual significance of QPOs (Huebner et al. 2021). Specifically, if the QPO is not present for the entire time series, frequency bins around the central QPO frequency are not statistically independent, which leads to an overestimate of the QPO significance. Furthermore, Auchère et al. (2016) have shown that detrending methods meant to eliminate the deterministic part of the time series cause artificial oscillatory behavior, which might be confused for real QPOs.

An alternative to frequency-domain methods is to analyze the data in the time domain, which has some advantages over periodogram-based methods. When calculating a periodogram, knowledge about heteroscedastic behavior, e.g., Poisson counting noise, is discarded and has to be inferred again from the periodogram. These measurement uncertainties can be included in time-domain analyses and thus better inform the results. Furthermore, using the time domain also obviates the need for signal windowing, as there is no requirement for the time series to be of infinite duration.

Gaussian processes (GPs; see Rasmussen & Williams 2006 for an in-depth introduction) describe a class of models that have enjoyed increasing popularity in astrophysics, due to their flexibility and recent improvements in computational performance. They enable us to parameterize both long-term deterministic trends and stochastic processes in the same model via mean functions and covariance functions. While they have been used for a range of different astrophysical data sets, e.g., as models for the cosmic microwave background (Bond & Efstathiou 1987; Bond et al. 1999; Wandelt & Hansen 2003), they have been particularly popular in time-series analysis. For example, Moore et al. (2016) propose to use GPs to account for uncertainty of gravitational-wave models, and D’Emilio et al. (2021) demonstrate how GPs can estimate the density of gravitational-wave posteriors. Relevant to our goal here of finding quasi-periodicities, there is some work demonstrating the viability of using GPs to analyze or search for periodicities in light curves, such as blazars (Covino et al. 2020) and quasars (Zhu & Thrane 2020).

The flexibility of the GP framework has made them a model of choice to account for structure in the data unimportant to the astrophysical problem of interest, e.g., modeling correlated instrumental noise (Gibson et al. 2012), modeling stellar variability in transit exoplanet searches (Grunblatt et al. 2015), accounting for pulsar timing residuals (van Haasteren & Vallisneri 2014), or as a model for stellar spectra in radial velocity searches (Czekala et al. 2017). However, for certain problems, they can also be employed as empirical models that can be related to astrophysical properties even when the covariance function itself bears no direct connection to the underlying physics. Examples include as a model for stellar spots, where a quasi-periodic covariance function and a GP can be used to infer stellar rotation (e.g., Angus et al. 2018), or in modeling stochastic variability in AGNs (Kelly et al. 2014). For some problems, it is possible to construct covariance functions that are directly related to the underlying physical properties of the observed source, as was recently done, for

example, by Luger et al. (2021) in an extension of the earlier work on stellar spots and by Lindberg et al. (2022) in modeling sparse asteroid light curves.

In this paper we discuss the use of GPs to perform searches for QPOs in the time domain. Focusing on transients specifically, we parameterize our model for the observed time series into a deterministic burst envelope, a stochastic component modeling rapid variability within the burst, and a candidate QPO, and we simultaneously infer details of these components through a parameterization of each via distinct mean and covariance functions. Although creating and evaluating general nonstationary QPO models is complicated, we show how we can make some relatively easy tweaks to stationary models to explore nonstationary structures. While this model is in principle applicable to any transient, our work here is motivated by high-energy transients in particular. For many of these transients, including GRBs, FRBs, and magnetar bursts, no generative physical model for their light curves exists. It is consequently challenging in these cases to define a null hypothesis or alternative model for QPO searches, motivating the choice of data-driven, empirical parameterizations in this paper.

Our approach here is situated somewhere between the second and third use case introduced above: as we will show below, the mean functions and covariance functions chosen to represent the bulk of the variability in all of our examples are strictly empirical, and while they generally fit the data well, they are not easily related to the astrophysical processes presumed to have generated the emission. On the other hand, the QPOs we hope to find in transient light curves generally are directly related to astrophysical interpretations, including starquakes in magnetars and magnetic reconnection in solar flares.

We show that we can use Bayesian inference to determine the properties of a QPO and to reliably perform model selection between QPOs and alternative models such as red noise. We lay out in detail what GPs are and motivate a kernel function that corresponds to QPOs in Section 2. In Section 3, we demonstrate on simulated data that both parameter estimation and model selection work reliably. Next, we demonstrate the viability on simulated light curves that Broomhall et al. (2019) produced for a comparative study between different QPO detection methods. We show in Section 5 that the method is easily transferable to many astrophysical data sets. Specifically, we reanalyze GRB090709A (Cenko et al. 2010; De Luca et al. 2010; Iwakiri et al. 2010), a magnetar burst from SGR 0501 (Rea et al. 2009; Huppenkothen et al. 2012), and a segment from the SGR 1806–20 giant flare (Hurley et al. 2005; Israel et al. 2005; Palmer et al. 2005; Strohmayer & Watts 2005, 2006; Watts & Strohmayer 2006; Huppenkothen et al. 2014a, 2014b; Miller et al. 2019). Beyond the phenomena discussed in this paper, there have been recent claims about QPOs in FRBs (Andersen et al. 2021; Pastor-Marazuela et al. 2022), where the method introduced here provides an alternative approach to look for signals. We discuss the results in Section 6.

2. Methods

In the following, we briefly introduce the core definitions of GPs, specifically in the one-dimensional case relevant to time-series analysis. For a more in-depth review, we refer the reader to Rasmussen & Williams (2006) and Foreman-Mackey et al. (2017).

In the following, we distinguish *deterministic* from *stochastic* data generation processes. In short, deterministic processes are those where the value of the time series at time t_i can be exactly forecast given a mathematical description of the data generation process. For example, given an amplitude, a phase, and a period, the value of a sinusoidal time series can be exactly forecast for any t_i . For stochastic processes, this is not true: here, the data generation process inherently involves a random element, such that future data points can only be described probabilistically. An example are random-walk-type processes such as Brownian motion. It is relevant here to distinguish *stationary* stochastic processes, where the mean and variance (weak stationarity) or the overall statistical properties (strong stationarity) of the process do not change with time, from *nonstationary* stochastic processes, where the mean and variance and the statistical properties are time dependent. Stochastic processes are common in astronomy, e.g., in accreting sources. The wide variety of burst and flare shapes observed, for example, in magnetar bursts and GRBs easily lends itself to the assumption that the underlying process generating these bursts and flares is stochastic.

2.1. Overview of Gaussian Processes

We can understand a GP as the sum of a deterministic process and a stochastic process. The deterministic process takes N time stamps $\{t_i\}_{i=1}^N$ —which can be evenly or unevenly spaced—and maps them to a flux y_i . For the purposes of this paper, we only consider single-band light curves, though we note that for many applications considering multiband light curves may improve results (VanderPlas & Ivezić 2015; Covino et al. 2022). The stochastic process assumes that the data points are draws from a multivariate Gaussian distribution with dimensionality N , where the variances and covariances describe the temporal relationships between data points. We can draw samples from the multivariate normal distribution similar to univariate distributions to use the GP as a generative model through a parameterization of the deterministic process, called the *mean function*, and the stochastic process, generally described by a parametric model called the *covariance function*. For forecasting purposes, the parameters of these functions are nuisance parameters. In the inference context of this paper, both the mean function and covariance function have parameters that are physically meaningful, and thus their posterior distributions reveal interesting information about the system under study.

The covariance function $k_\alpha(t_i, t_j)$ with a vector of parameters⁹ α defines a covariance matrix $K(\alpha)$ of size $N \times N$. We write the mean function as $\mu_\theta(t)$, with parameter vector θ .

We define the coordinate vector \mathbf{t} (in our case, the coordinate of interest is time, but in practice, it need not be) and data vector \mathbf{y} with N entries each. The GP log-likelihood is thus

$$\ln L(\theta, \alpha) = -\frac{1}{2} \mathbf{r}_\theta^T K_\alpha^{-1} \mathbf{r}_\theta - \frac{1}{2} \ln \det K_\alpha - \frac{N}{2} \ln(2\pi), \quad (1)$$

where

$$\mathbf{r}_\theta = \mathbf{y} - \mu_\theta(\mathbf{t}) \quad (2)$$

⁹ In the context of GP regression, the literature refers to the parameters α as “hyperparameters,” which should not be confused with the hyperparameters used in hierarchical inference.

is the residual vector. The covariance matrix elements are given by the kernel function $[K_\alpha]_{nm} = \sigma_n^2 \delta_{nm} + k_\alpha(t_n, t_m)$, where σ_n^2 are the variances due to white noise and δ_{nm} is the Kronecker delta. In the simplest case of $k_\alpha(t_n, t_m) = 0$, the likelihood function becomes a simple Gaussian likelihood that one would typically employ in least-squares regression. The fact that we can include Gaussian measurement errors σ_n , which manifest themselves as a white-noise floor in periodograms but can otherwise not be taken into account, means that we correctly treat heteroscedastic data. In the above definition of the covariance matrix, we have assumed a stationary covariance function, such that the kernel only depends on the *time difference* τ between different y_i , i.e., $k(t_n, t_m) = k(|t_n - t_m|) = k(\tau)$. We use different covariance functions to model broadband variability (also known as *red noise*), as well as QPOs. Nonstationary covariance functions exist but are computationally challenging to evaluate. For the majority of this work, we assume that a significant fraction of the variability in fast transients can accurately be described by a stationary stochastic process, and nonstationarity enters the model primarily through the mean function.

One of the main drawbacks of GPs is that, in general, the likelihood takes $\mathcal{O}(N^3)$ steps to evaluate (Rasmussen & Williams 2006). There are numerous sophisticated approaches to reduce this complexity to a more manageable level (Flaxman et al. 2015; Wilson & Nickisch 2015; Gardner et al. 2018), including GPU acceleration (Delbridge et al. 2019). Most suitable for our purposes is the CELERITE software package, which solves performance issues by restricting itself to one-dimensional problems and the class of sums and products of complex exponential kernel functions, which reduces the complexity to $\mathcal{O}(NJ^2)$ (Foreman-Mackey et al. 2017), J being the number of exponential terms that form the kernel. CELERITE achieves this by exploiting the semiseparable structure of the covariance matrices, which allows for the use of a fast solver for the Cholesky factorization (Foreman-Mackey et al. 2017). We investigate in the following section why the class of exponential kernel functions is sufficient for our purposes.

2.2. Kernel Functions

There is an exhaustive literature on popular kernel functions, but it is not always clear which kernel function is most suitable to use and how to interpret their hyperparameters. In fact, for many applications of GPs, this is not even the goal. Instead, one may be interested in a GP’s predictive power to interpolate and extrapolate data, which can be tested and tuned on training sets. However, we are interested in interpreting the kernel function and associating it with an underlying physical process. For example, we want to infer a posterior distribution of the frequency of a QPO.

We start by defining a kernel describing a periodic oscillation

$$k_{\text{po}}(\tau) = a \cos(2\pi f\tau), \quad (3)$$

where a is the amplitude of the oscillation and f is its frequency. This kernel corresponds to a perfect harmonic oscillation; its representation in the power spectrum is a Dirac delta function peaking at f . In order to make this kernel quasi-periodic, we add

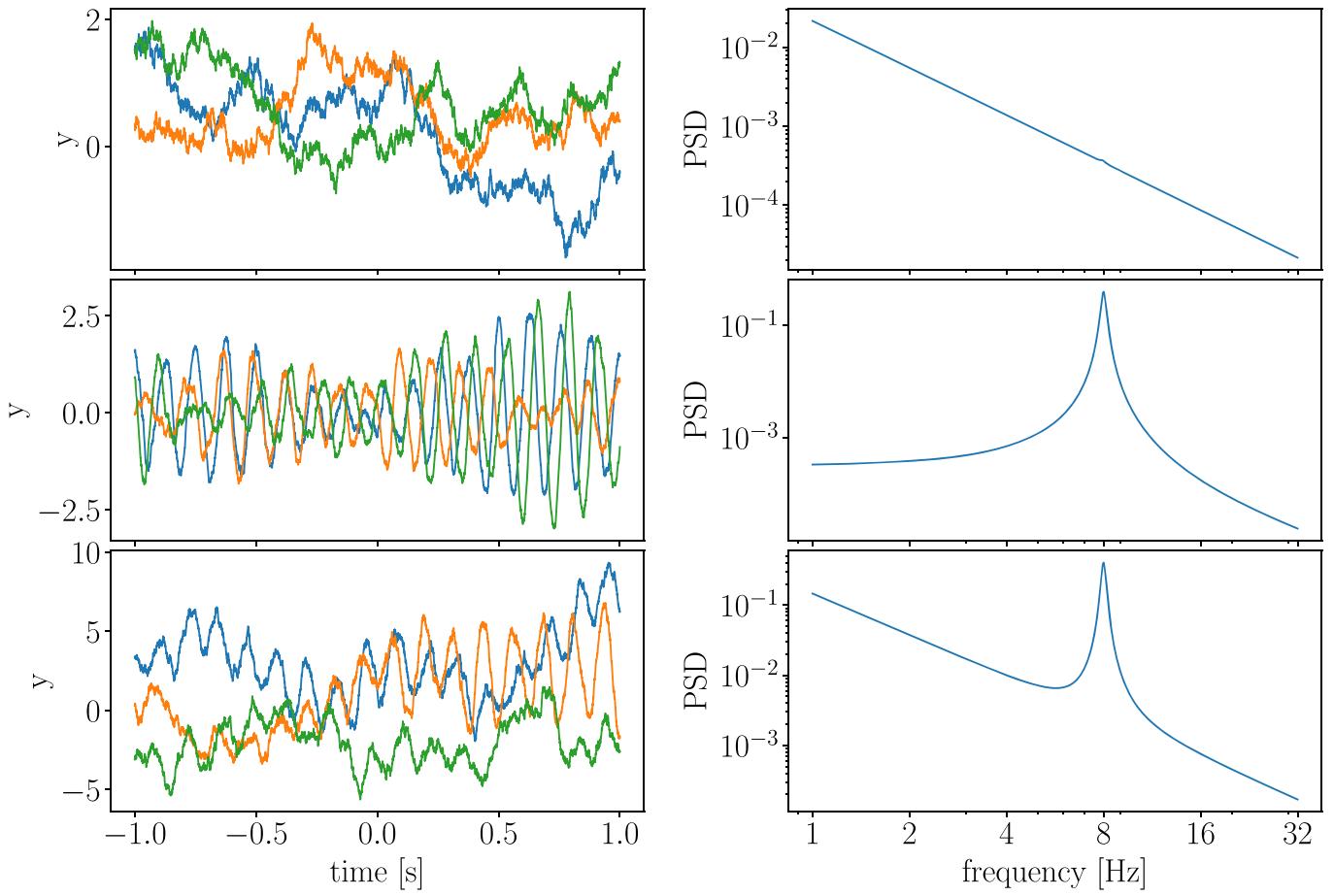


Figure 1. Realizations of GPs with a red noise k_m (top left), quasi-periodic k_{qpo} (middle left), and combined red noise and quasi-periodic $k_{\text{qpo}+m}$ kernels (bottom left). The panels on the right are the corresponding power spectra to the time series on the left. We set $a_{\text{qpo}} = a_m = c_{\text{qpo}} = c_m = 1$ and $f_{\text{qpo}} = 8$ in the first and second panel, and we use the same parameters except for $a_m = \exp(2)$ in the third panel.

another factor to account for variations in amplitude over time,

$$k_{\text{qpo}}(\tau) = \exp(-c\tau)k_{\text{po}}(\tau), \quad (4)$$

where c is the inverse of the decay time of the QPO. The c parameter indicates that if we generated a time series using this kernel, the QPO would be excited or damped on a typical timescale of $1/c$. In the frequency domain, this kernel corresponds to a Lorentzian function with peak frequency f and FWHM c . Since GPs are also generative models, we plot some realizations of the k_{qpo} kernel in the top left panel of Figure 1.

In practice, many astrophysical systems will display both red noise and QPOs at the same time. In the context of CELERITE, we can model red noise as a simple exponential

$$k_m(\tau) = a \exp(-c\tau). \quad (5)$$

Physically, this kernel corresponds to a damped random walk (Ornstein-Uhlenbeck process), and its power spectrum is an f^{-2} power law. While this kernel provides a convenient and fast prescription of a correlated noise process, we caution the reader that this describes *one* specific noise process, which may not be appropriate for a given application. Wherever possible, prior knowledge of the overall process that may have generated the data (without the QPO) should be used to design a covariance function applicable to that problem. Since $k_m(\tau)$ is not mean-square differentiable, the model also implies that

corresponding time series are not differentiable (Rasmussen & Williams 2006). We show some realizations of k_m in the top left panel of Figure 1.

In order to obtain a kernel that describes the QPO and the red-noise process, we simply add both kernels

$$k_{\text{qpo}+m}(\tau) = k_{\text{qpo}}(\tau) + k_m(\tau). \quad (6)$$

This kernel is somewhat different from the QPO kernel proposed in Foreman-Mackey et al. (2017),

$$k(\tau) = \frac{a}{2+b} \exp(-c\tau) [\cos(2\pi f\tau) + (1+b)], \quad (7)$$

which also models both QPO and red-noise features but makes them share the c parameter. Since we do not find that $c_{\text{qpo}} = c_m$ in general, we use Equation (6) as our model for the general case and Equation (5) for the case in which we only encounter red noise.

There are alternative ways to describe periodic behavior outside the domain of CELERITE-compatible kernel functions. Rasmussen & Williams (2006) used the following kernel to model seasonal changes in atmospheric CO₂ levels:

$$k(\tau) = a \exp \left[-\frac{\tau^2}{2\ell^2} - \Gamma \sin^2(\pi f\tau) \right]. \quad (8)$$

Angus et al. (2018) demonstrated that this kernel can be used to infer stellar rotation periods. This kernel combines a squared

exponential with an oscillatory term such that a parameter Γ controls the amount of covariance between two points that are roughly one period away from each other. For high values of Γ , only points exactly integer multiples of a period away from each other have high covariance. Similarly to k_{qpo} , Equation (8) is quasi-periodic in that it is damped, though not exponentially but with the squared exponential. This way, arbitrary curves with a repeating shape can be modeled, though a QPO kernel as in Equation (6) has shown to be sufficient to correctly infer the frequency for arbitrary oscillatory curves (Foreman-Mackey et al. 2017). Furthermore, while Equation (8) is practical to fit nonsinusoidal curves, it may be hard to interpret the underlying physics from the inferred parameters and comes with the high computational cost of general GP kernels. Alternatively, Equation (8) could also be approximated by a Fourier series since the CELERITE kernel family forms a Fourier basis. However, this approach would also incur far higher computational costs since the CELERITE likelihood complexity scales quadratically with the number of terms required to build the kernel, and the additional Fourier coefficients would drastically increase the parameter space.

There are also alternative ways to define aperiodic kernels. Within the support of CELERITE models, Foreman-Mackey et al. (2017) proposed using a critically damped stochastic harmonic oscillator

$$k(\tau) = S_0 \omega_0 e^{-\frac{1}{\sqrt{2}}\omega_0\tau} \cos\left(\frac{\omega_0\tau}{\sqrt{2}} - \frac{\pi}{4}\right), \quad (9)$$

which is commonly used to model background granulation noise in asteroseismic and helioseismic oscillations. The stochastic harmonic oscillator is once mean-square differentiable and thus yields smoother time-series data than the red-noise kernel. Outside of the class of sums and products of complex exponentials that CELERITE supports, the squared exponential, Matern-3/2, and rational square kernel functions are employed in many contexts. Recently, some of these kernels have also been applied to search for oscillations in blazar light curves (Covino et al. 2020). However, these kernels are difficult to motivate physically, and the higher computational complexity may restrict their use to maximum likelihood estimates rather than Bayesian evidence calculations. We stick to the red-noise kernel for our analyses, as we consider this to be a basic framework that may describe many physical systems to a first degree.

2.3. Mean Functions

Choosing a mean model is similarly tricky to choosing a kernel function if there is no physically motivated model for the underlying background trend. The simplest mean model takes the (weighted) average of all recorded data points. However, this model implicitly assumes that there are no trends in the data and thus all variability in the light curve was generated by a stochastic process that can be described by the chosen covariance function. If a significant background trend is present, this assumption quickly breaks down and leads to incorrect inferences about the variability in the light curve. It is thus not a suitable way to approach the problem in most cases.

Another approach is to use filter methods such as the boxcar filter or the Savitzky–Golay filter (Savitzky & Golay 1964) and subtract the filtered light curve from the original light curve. However, using filtering methods is prone to cause artificial periodicities (Auchère et al. 2016). Some problems feature a relatively simple background trend that one could model as a linear trend or another low-order polynomial instead of a filter. For some problems, physical or phenomenological models may exist to describe the overall trend of the light curve, such as the “fast-rise exponential decay” (FRED) model for GRB light curves.

For general problems, one can employ methods such as shapelet fitting, in which one casts the problem as a sum of orthonormal basis functions. Alternatively, for burst-like light curves, it is often possible to identify several base flare shapes (e.g., (skewed) Gaussians, FREDs; Huppenkothen et al. 2015) that one can use to model the overall shape of the burst.

Here we use a number of very simple phenomenological prescriptions as flare mean models, and we urge the reader to consider physically meaningful options for mean functions in their specific application. As the most basic models we use skewed Gaussians and skewed exponentials, i.e.,

$$\mu_{\text{GAUSS}}(t; A, t_0, \sigma_1, \sigma_2) = \begin{cases} A \exp\left(-\frac{(t - t_0)^2}{2\sigma_1^2}\right) & \text{if } t \leq t_0 \\ A \exp\left(-\frac{(t - t_0)^2}{2\sigma_2^2}\right) & \text{if } t > t_0 \end{cases} \quad (10)$$

and

$$\mu_{\text{EXP}}(t; A, t_0, \sigma_1, \sigma_2) = \begin{cases} A \exp\left(-\frac{t - t_0}{\sigma_1}\right) & \text{if } t \leq t_0 \\ A \exp\left(\frac{t - t_0}{\sigma_2}\right) & \text{if } t > t_0. \end{cases} \quad (11)$$

There are a number of other approaches to define FRED models. We focus on the FRED model defined in Norris et al. (1996), which can assume a wide variety of shapes:

$$\begin{aligned} \mu_{\text{FRED}}(t; A, t_0, \psi, \Delta) \\ = A \exp\left[-\psi\left(\frac{t + \Delta}{t_0} + \frac{t_0}{t + \Delta}\right)\right] \exp(2\psi) \end{aligned} \quad (12)$$

for $t < 0$. This model’s maximum coincides with $t = t_0$, and we can interpret Δ as an “offset” parameter that allows us to fit the full range of flare shapes with arbitrary offsets from $t = 0$. Finally, ψ serves as a symmetry parameter, with the curve becoming increasingly symmetrical for large values of ψ . The additional factor $\exp(2\psi)$ is optional in the model definition and serves as a normalization factor such that $\mu(t = t_0; A, t_0, \psi, 0) = A$, which eases the definition of the amplitude prior.

We display the mean models in Figure 2. As we show later, all models we have introduced here reasonably fit general flare shapes. Performing Bayesian inference with these models thus naturally yields comparable evidences. In this paper, we consider models that contain a single instance of one of these mean models, but we also consider mean functions that are composed of a superposition of multiple components, e.g., a model with three skew-Gaussian components, each with their

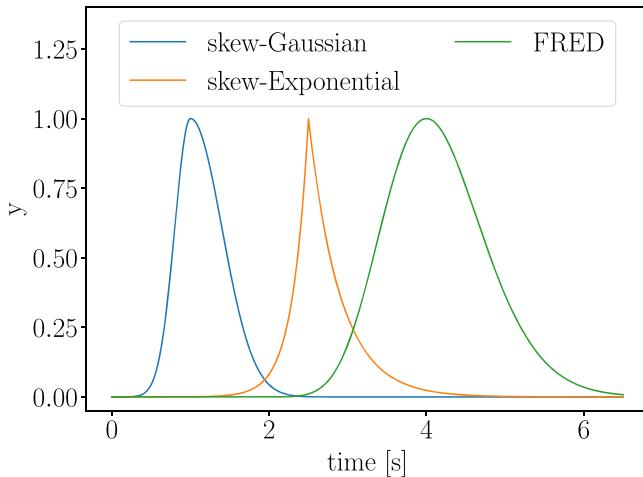


Figure 2. Example plots of the three mean models we use throughout.

own parameters. For general problems, it is prudent to run the analysis with multiple mean models and select the one that yields the highest evidence.

2.4. Toward Nonstationary Models

We have so far laid out how GPs resolve the issues with heteroscedastic uncertainties and deterministic trends that lead to biases in Fourier-based analyses. The other issue we face is whether the QPO and noise process are nonstationary. Bursts and flares are inherently nonstationary processes: it is possible that the properties of the noise process change as a function of time within a burst. Similarly, it is possible that the QPO is only present for part of the burst, as is the case in neutron starburst oscillations (Watts 2012).

If we do not select the data segment we want to analyze carefully, we are likely to incur similar nonstationarity biases to periodogram-based analyses as Huebner et al. (2021) laid out. However, we do not know *a priori* whether the variability processes in the data are stationary or not, or whether a QPO may exist over the full length of the light curve or only part of it. Ideally, we would like to infer these properties, rather than assume stationarity over the full time series.

As a simple example, we consider a system with constant X-ray emission characterized by white noise, which flares for some time and then goes back to its regular constant emission. We describe this scenario by first defining a continuous mean function to be present through the entire time series (here a constant, though, in practice, it could be more complex). Additional variability on top of that shape is present in three distinct segments, where the stochastic variability in the first and last segments is described solely by white noise, and the middle segment, which also contains both a QPO and red noise. We show how we can implement this basic, nonstationary process and use it to infer the QPO’s start and end.

By breaking the time series into three segments, we can model a basic, nonstationary process without resorting to a more general likelihood solver. Instead, we split the likelihood into two parts, one for the two disjoint, white-noise segments on either end and the GP in the center. The covariance matrix can then be decomposed into two submatrices correspondingly. We obtain one “outer” submatrix by deleting all rows and columns that contain off-diagonal entries, and one covariant “inner” submatrix as the complement of the outer submatrix.

Since the inner covariant submatrix forms a block, we can calculate the inverse and the determinant independent from the remaining entries. The same operations are trivial for the outer submatrix. We rewrite Equation (1) to describe this explicitly:

$$\begin{aligned} \ln L(\theta, \alpha) = & -\frac{1}{2} \sum_{n_{\text{out}}} \frac{r_{n_{\text{out}}}^2}{\sigma_n^2} - \sum_{n_{\text{out}}} \ln \sigma_{n_{\text{out}}} - \frac{N_{\text{out}}}{2} \ln(2\pi) \\ & - \frac{1}{2} r_{\text{in},\theta}^T K_{\text{in},\alpha}^{-1} r_{\text{in},\theta} - \frac{1}{2} \ln \det K_{\text{in},\alpha} - \frac{N_{\text{in}}}{2} \ln(2\pi) \end{aligned} \quad (13)$$

where the likelihood coming from the outer noncovariant submatrix is in the first line and the inner covariant submatrix is in the second line. Here n_{out} are the indices of the outer N_{out} noncovariant row/columns, $K_{\text{in},\alpha}$ is the inner covariant submatrix, and r_{in} are the residuals of the N_{in} inner rows/columns. We introduce the t_{start} and t_{end} parameters to our model to describe the transition points between the segments, to be inferred together with the remaining parameters.

2.5. Assumptions and Limits of GPs

As with all other methods, the reliable detection of QPOs in time series crucially depends on whether the assumptions of the method and of the model proposed to generate the data are reasonable ones. For many of the examples we showcase in this paper, no realistic physical model exists capable of generating realistic simulated data. Therefore, QPO detection necessarily relies on imperfect, phenomenological prescriptions to describe the data generation process with or without QPO. Here we assume that the observed time series can be described as the combination of a simple flare shape and a stochastic process. This is a more realistic extension of the assumptions that all variability can be described by a stationary stochastic process that previous approaches using Fourier methods have made. As our understanding of these phenomena improves, better assumptions about the physical processes that generate these bursts and flares will significantly improve QPO detections.

One fundamental assumption we make is that QPOs are an additive process, i.e., the QPO does not depend on the value of the mean function at any point. This assumption is likely not a perfect model. At the very least, we expect no QPOs long before and after a transient. This observation implies some connection between the mean function and the GP, which is challenging to model using our framework.

The QPO model we propose is also inherently stationary, as this is required for the fast Cholesky solver implemented in the CELERITE package, and we can only model a narrowly defined set of nonstationary time series as we have shown in Section 2.4. QPOs that drift in frequency over long timescales cannot be adequately modeled and require us to use a different framework of GP solvers. Once frequency shifts become too large, our QPO kernel is sure to fail. We can detect shifting frequencies by splitting the light curve into multiple segments that we analyze independently and compare the QPO frequency posteriors, or via hierarchical Bayesian inference. We would have to evaluate a more expensive nonstationary QPO model for a more rigorous analysis.

Many of the phenomena considered here are bright, energetic flares observed in X-rays. Observatories typically record astrophysical X-ray data by counting the number of photons that arrive at a detector in a specific time interval, a

process by its nature Poissonian. GP modeling, however, inherently assumes that the underlying data are Gaussian. While Poisson counting data are Gaussian to a good approximation if there are sufficient counts per bin, GPs do not correctly model data with lower count rates. There are some approaches to deal with photon-counting data with low count rates. The easiest way is to apply a variance-stabilizing transform, which makes the data approximately Gaussian with $\sigma = 1$ (Anscombe 1948; Bar-Lev & Enis 1990). However, these transformations are still far off for bins with zero or one photon. Other methods like a sigmoidal Gaussian–Cox process treat Poisson data effectively as Gaussian by converting the data with a sigmoid function (Adams et al. 2009; Flaxman et al. 2015), but using these methods makes it harder to interpret the parameters because the sigmoid function also transforms them.

Real data are often more complicated than a simple Poisson process and may feature additional effects such as dead time or double-counting of photons. We assume that dead time, i.e., the property of detectors that they cannot detect a second photon for a short period after they counted the first photon, does not play a significant role in our data. Dead-time effects are particularly a problem for strongly flaring sources. Modeling dead time in practice is complex and requires sophisticated techniques such as simulation-based inference (Huppenkothen & Bachetti 2021).

2.6. Bayesian Inference with Gaussian Processes

Bayesian inference is a statistical paradigm for parameter estimation and model selection. We used Bayesian inference with the calculation of Bayes factors in Huebner et al. (2021) to characterize model preference and thereby the significance of QPOs within time-series data, and we apply the same approach here. To understand the meaning of Bayes factors, we consider Bayes's theorem

$$p(\theta|d, M) = \frac{\pi(\theta|M)L(d|\theta, M)}{Z(d|M)}, \quad (14)$$

where θ are both the mean and kernel parameters; d are the data, i.e., the time series in our case; $p(\theta|d, M)$ is the posterior probability of the parameters; $\pi(\theta|M)$ is the prior probability of the parameters; $L(d|\theta, M)$ is the likelihood of the data given the parameters; and $Z(d|M)$ is the evidence, or fully marginalized likelihood. All these probabilities are conditioned on the choice of kernel and mean model, jointly encoded in the placeholder variable M , which we want to evaluate. The evidence describes an overall probability of the given model producing the data and can be calculated by rearranging and integrating Bayes's theorem

$$Z(d|M) = \int \pi(\theta)L(d|\theta, M)d\theta. \quad (15)$$

Equation (15) is expensive to solve using simple grid methods, which is why we have to employ sophisticated algorithms such as nested sampling in practice (Skilling 2004, 2006).

Though the evidence itself carries no straightforward intrinsic meaning as a normalization factor, taking the ratio between two yields the Bayes factor

$$BF = \frac{Z(d|M_1)}{Z(d|M_2)}, \quad (16)$$

where M_1 and M_2 are the different models that yield the respective evidences. The Bayes factor measures the odds of the underlying data being produced by either model, assuming that both models are equally likely to be correct, though it does not measure whether the model itself is a good fit to the data. Model checking for the goodness of fit of an individual model on the data can be performed, for example, through posterior predictive checks (e.g., Gelman et al. 2013; Blei 2014; Lucy 2016, 2018). Because Bayes factors are sensitive to the prior volume, care should be taken when choosing priors and interpreting the results.

Throughout this paper we use the BILBY package to implement Bayesian inference (Ashton et al. 2019a; Romero-Shaw et al. 2020). We wrote a small likelihood interface for CELERITE and BILBY, which is now part of the main BILBY package. All software used to produce the results in this paper can be found in its repository.¹⁰

Bayes factor calculations for model selection of GPs usually have not been considered viable (notable exceptions are Covino et al. 2020; Zhu & Thrane 2020). One can compute a maximum likelihood estimate much faster and derive alternative measures, such as the Akaike or Bayesian information criterion. However, these measures are less reliable if the posterior distribution is not Gaussian or features multiple modes. When dealing with CELERITE models, performing the necessary likelihood evaluations for complete posterior and evidence calculations is relatively fast. We find that nested sampling can be performed in $\mathcal{O}(\text{minutes})$ using CELERITE for a light curve with a few hundred t_i when using the k_{qpo} GP kernel and a skewed Gaussian mean model. Furthermore, even beyond the class of CELERITE models, full Bayesian inference may be possible on many data sets with the use of efficient GP evaluation algorithms (Flaxman et al. 2015; Wilson & Nickisch 2015; Gardner et al. 2018; Delbridge et al. 2019) or the use of massively parallel inference (Smith et al. 2020).

2.7. Priors

We must choose priors for the parameters of the covariance function α and of the mean function θ carefully. Here we largely choose uninformative priors not tied to a specific application, but we urge readers to consider informative priors for their specific application carefully. Making priors informative is important because the prior range impacts the evidence calculation: an overly large prior range would suppress the evidence and disfavor the model. However, in problems where knowledge about the underlying physical processes is sparse, setting informed priors is often not possible or impractical. For example, the possible amplitudes a depend on both the efficiency of the instrument and the intrinsic brightness of the object we observe. Thus, we set prior bounds for amplitudes based on the data.

Generally, we employ priors uniform in logarithm for scale parameters of the covariance function such as the amplitude a , inverse decay time c , and frequency f since these prior spaces span several orders of magnitude.

In theory, we can set the prior range of a_m and a_{qpo} arbitrarily wide. However, in some circumstances this prior choice causes numerical issues. For example, if there is no QPO present in the data, the a_{qpo} posterior will have an upper

¹⁰ <https://github.com/MoritzThomasHuebner/QPOEstimation>

limit below which it is identical to the prior distribution. Thus, we set the lower limit of both the a_{rn} and a_{qpo} prior to be the smallest σ_n , meaning that we assume that the QPO/red-noise process has to be at least as strong as the white noise. We set the upper limit on a_{rn} and a_{qpo} to be twice the difference between the highest and lowest value within the time series.

Assuming that the data points are equally spaced, the frequency parameter f should not be smaller than than $1/T$, with T being the length of the data segment and not being greater than the Nyquist frequency, which is half the sampling rate. The Nyquist frequency for unevenly sampled data is $1/2p$, where p is the largest value such that each time value t_i can be written as $t_i = t_0 + n_i p$, with n_i being an integer (Eyer & Bartholdi 1999; VanderPlas 2018). We note that this upper limit might lead to QPO searches up to very high frequencies and stress the importance of setting physically motivated upper limits for the QPO frequency in practical applications.

The upper limit of c_{rn} should be not greater than the sampling frequency $f_s = 1/\Delta t$ of the data because otherwise the kernel will decay to less than $1/e$ within Δt . Beyond this point, the red-noise model is not functionally different from a white-noise model. If we consider unevenly sampled data, we can use the pseudo-Nyquist frequency defined above as an upper limit to reflect the prior choice for the frequency f . As for the frequency prior, the upper limit of c_{rn} should reflect physically meaningful boundaries. We set the lower limit of c_{rn} to $1/T$ to ensure that the red-noise process occurs on a timescale at a maximum still comparable to T . While lower values of c_{rn} may have posterior support if permitted, they imply a constant offset between the mean function and the data that is not physically motivated.

Unlike with c_{rn} , a very low c_{qpo} cannot cause a constant offset since the k_{qpo} averages to zero over long timescales. Thus, c_{qpo} can, in principle, take on infinitely low values. We set the minimum of c_{qpo} to be $1/(10T)$, which means that a k_{qpo} has to decrease by $\approx 10\%$ throughout the data segment, as we cannot functionally distinguish lower values. For the upper limit of c_{qpo} we need to ensure that k_{qpo} is periodic. Thus, we demand that k_{qpo} does not decay to less than $1/e$ over one period, which implies $c_{\text{qpo}} \leq f_{\text{qpo}}$. We implement this relation in practice using the `Constraint` prior feature in BILBY.

Finally, we also consider the priors of our analytic mean flare shapes. Again, these generally depend on the details of the underlying physical process and our knowledge of it. For the bursts and flares considered in this paper, many of our priors are derived directly from properties of the observations, under the assumption that the properties of transient detection are often driven by the source's behavior. For example, under the assumption that the instrument observed the entire flare down to some background, we can choose a uniform prior for the parameter describing the peak of the mean function limited to the start and end of the observation. The amplitude priors are uniform in logarithm between 10% and 200% of the difference between minimum and maximum of the data. The prior on the width parameter of our flares depends on the model function used for the mean model. In general, we aim for the width parameter to have a uniform-in-logarithm prior that spans roughly from the time difference between two data points to the entire time range of the data.

For mean models defined as a superposition of multiple simple shapes, the prior for the peak time requires careful thought because the parameters of each model component are degenerate under permutation of the order of these components. If we did not

break this degeneracy, the parameter would be much harder to sample since n mean model components allow for $n!$ degenerate permutations. Assuming n components making up the mean model throughout the interval, we can use a result from order statistics to derive appropriate priors. We label the flares in order from 0 to $n-1$ indexed by k . The $t_{0,k}$ prior is distributed according to a conditional beta distribution with $\alpha=1$ and $\beta=n-k$ between the $t_{0,k-1}$ flare and the end of the interval (Gentle 2009). Putting this together, we obtain

$$\pi(t_{0,k}|t_{0,k-1}) = (n-k) \left(1 - \frac{t_{0,k} - t_{0,k-1}}{t_{\max} - t_{0,k-1}}\right)^{n-k-1}, \quad (17)$$

where t_{\max} is the time of the last element of the light curve. For the special case of $k=0$ the prior is not conditional and we have instead

$$\pi(t_{0,k=0}) = n \left(1 - \frac{t_{0,k=0} - t_{\min}}{t_{\max} - t_{\min}}\right)^{n-1}, \quad (18)$$

where t_{\min} is the first element of the light curve.

The priors on the parameters of nonstationary extension of the CELERITE model, t_{start} and t_{end} , are chosen in the same way as in Equations (17) and (18), but with $n=2$.

As a final note, the priors here are chosen to be empirical and relatively uninformative, with the goal of showcasing the procedure rather than enabling astrophysical inference on a specific problem. We present the priors used throughout this paper for both simulations and analysis of astronomical data in Table (1), and note deviations from those choices where relevant. In practice, careful design of the priors for any given physics problem is important in the context of *population inference*, where the priors described above would likely lead to significant biases.

3. Simulated Data

In this section, we look at different configurations of the models we described earlier and validate them on simulated observations. We examine the QPO kernel and our red-noise kernel with different parameters. We also investigate what happens when we infer parameters with the wrong model, e.g., if we create red-noise data and perform inference with $k_{\text{qpo+rn}}$.

We generate GP data using a set of parameters in two steps. First, we obtain the deterministic part of the data by evaluating the mean function. Second, we produce the stochastic process arising from the covariance between the data points. We create the stochastic process data by drawing a sample from the multivariate normal distribution associated with this covariance matrix. Finally, we add the deterministic and stochastic parts to create the overall time series.

3.1. Percentile–Percentile Analysis

Before we begin applying our methods, it is important to establish their validity on simulated data sets. Specifically, it is hard to know a priori how to tune sampling settings within BILBY and the nested sampling package we use here, DYNESTY (Speagle 2020), to obtain adequate results. Though gravitational-wave researchers frequently use BILBY and have established settings (Romero-Shaw et al. 2020), the data sets and models we use have different dimensionality and posterior distributions that may be more difficult to sample from. Specifically, we want to establish that the posterior samples we draw are representative of the true posterior distribution.

Table 1
General Prescriptions for the Priors Used in This Paper, Largely Based on the Properties of the Data

Parameter	Prior Class	Minimum	Maximum
a_m	LogUniform	$\min \sigma_n$	$2\Delta y_{\max}$
c_m	LogUniform	$1/T$	$2f_{\text{Nyquist}}$
a_{qpo}	LogUniform	$\min \sigma_n$	$2\Delta y_{\max}$
c_{qpo}	LogUniform	$1/(10T)$	f_{qpo}
f_{qpo}	LogUniform	$1/T$	f_{Nyquist}
A_m	LogUniform	$0.1\Delta y_{\max}$	$2\Delta y_{\max}$
$t_{m,0}$	Uniform	t_{start}	t_{end}
$\sigma_{m,1}$	LogUniform	$\min(t_{i+1} - t_i)$	T
$\sigma_{m,2}$	LogUniform	$\min(t_{i+1} - t_i)$	T
t_{start}	Equations (17), (18)		
t_{end}	Equations (17), (18)		

Note. These priors are generic and uninformative and make use of the assumption that the data processing encodes relevant information about the transient's properties (e.g., its duration). Note that $\Delta y_{\max} = (\max y - \min y)$. Log-uniform priors were chosen for scale parameters spanning multiple orders of magnitude, uniform parameters for all other parameters. The parameters for the mean function (those with subscript m) should be considered placeholders for the individual parameterizations used in the remainder of the paper, but they provide a general overview of the priors implemented as part of the injection studies in Section 3 and when using the model to analyze astrophysical observations in Section 5. Note also that additional constraints are implemented on the peak time $t_{m,0}$ as described in Section 2.7 in order to account for permutation degeneracies when the mean model consists of multiple components.

Table 2
Priors for PP Test

Parameter	Prior Class	Minimum	Maximum
a_m	LogUniform	$\exp(-1)$	$\exp(1)$
c_m	LogUniform	$\exp(-1) s^{-1}$	$\exp(1) s^{-1}$
a_{qpo}	LogUniform	$\exp(-1)$	$\exp(1)$
c_{qpo}	LogUniform	$\exp(-1) s^{-1}$	$\exp(1) s^{-1}$
f_{qpo}	LogUniform	1 Hz	64 Hz
A	LogUniform	10	100
t_0	Uniform	0 s	1 s
σ_1	LogUniform	0.1 s	1 s
σ_2	LogUniform	0.1 s	1 s

Note. The prior is used to randomly draw parameters for simulated data sets and during the inference process. There is also a prior constraint such that $c_{\text{qpo}} < f_{\text{qpo}}$, which we introduced in Section 2.6.

If the inference process is unbiased, then the actual value of the parameters has to be in the x th-percentile of the posterior distribution x percent of the time. We can use this percentile-percentile (PP) analysis to tune the sampling settings based on simulated data before deploying the analysis on real data (Cook et al. 2006). The PP analysis is a particular case of a quantile-quantile analysis in that both considered distributions are uniform between 0 and 1. PP tests fail if the settings are insufficient to sample the posterior without bias, e.g., if we use too few live points or if some of the parameters are hard to sample. Aside from difficulties in sampling, PP tests should pass when using the same model and priors for data creation and inference. Concretely, the PP test finds a p -value for each parameter using a Kolmogorov–Smirnov test (Massey 1951) to check whether the fraction of events in a particular credible interval is drawn from a uniform distribution. The PP test

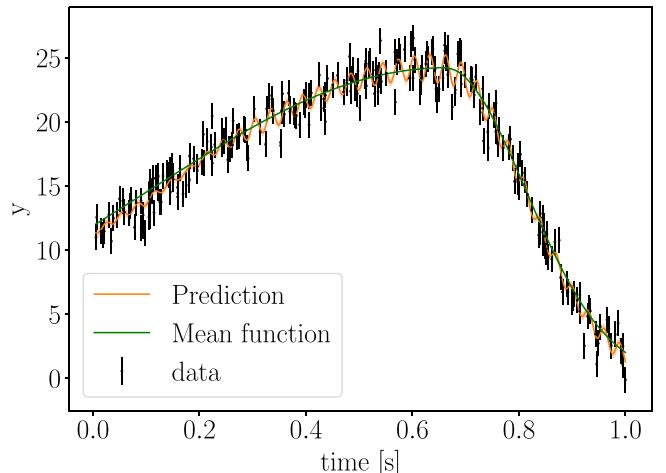


Figure 3. Example draw from the prior distributions given by the parameters in Table 2. We show the data in black and the mean function in green. The prediction curve with 68% confidence bands in orange shows, based on the underlying parameters, an estimate of the y -values between the data points.

combines these individual p -values to a combined p -value. Conventionally, $p < 0.05$ indicates that the fraction of events is not uniformly distributed to the 2σ level.

We perform the PP test using the $k_{\text{qpo}+m}$ model and a single skewed Gaussian as an example mean model. We create 100 simulated 1 s long data sets sampled with 256 time bins t_i that are randomly uniform distributed in time and using parameters randomly drawn from the priors in Table 2. We display one random draw in Figure 3. Choosing randomly spaced t_i does not require any additional effort since the prior boundaries are fixed and has the benefit of validating the method for a broader set of possible problems.

Figure 4 shows the PP analysis using DYNESTY's random walk sampling with 1500 live points and BILBY default settings. This result shows that these settings are sufficient to sample the parameter space without significant biases in the posterior. The overall p -value of 0.1528 indicates that the observed deviations are consistent with randomness, though the individual relatively low values for f_{qpo} and a_m indicate that these parameters may be harder to sample and may benefit from more finely tuned settings. We find similar results using DYNESTY's random slice sampling, which is substantially faster but generally produces a worse representation of the posterior (Romero-Shaw et al. 2020).

3.2. Model Selection

We want to understand how stable and reliable Bayes factors are in QPO detection given multiple light curves generated from a model with the same parameters, since noise realizations of the same parameters can be vastly different. For this study, we create the data with a symmetric Gaussian mean function. We create one set of simulated light curves using k_m and two sets using $k_{\text{qpo}+m}$ with different QPO amplitudes. We list the parameters in Table 3. The first QPO amplitude is relatively low compared to the noise in the data, whereas the higher QPO amplitude corresponds to a fourfold increase over the lower one. We use both kernel functions plus the skewed Gaussian mean model to carry out model selection between a model with a QPO and a model without a QPO for all light curves. We

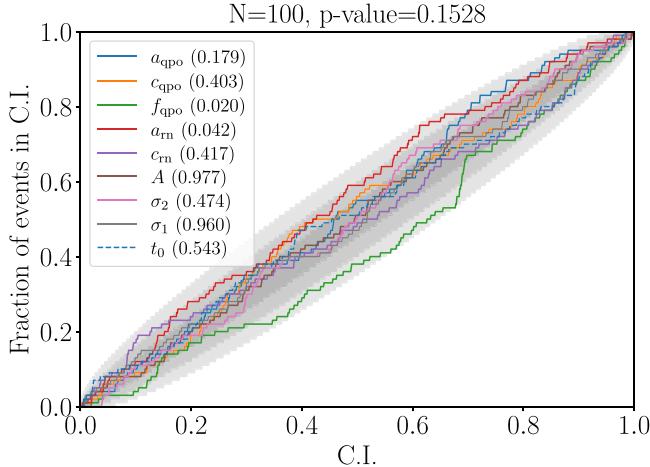


Figure 4. The PP plot shows that our sampling methods are not biased. We obtain the plot using the respective function in BILBY (Romero-Shaw et al. 2020). We show the confidence interval (CI) on the horizontal axis and what fraction of events have the true value within the CI on the vertical axis. The gray bands indicate the 1σ , 2σ , and 3σ levels. The overall p -value of 0.1528 indicates that the observed deviations are consistent with randomness, though the individual relatively low values for f_{qpo} and a_{rn} shown in parentheses in the legend indicate that these parameters may be harder to sample and may benefit from more finely tuned settings.

Table 3
Simulation Parameters for Model Selection Study

Parameter	k_{rn} Values	$k_{\text{qpo+rn}}$ Values
a_{rn}	$\exp(1)$	$\exp(1)$
c_{rn}	$\exp(1) \text{ s}^{-1}$	$\exp(1) \text{ s}^{-1}$
a_{qpo}	...	$\exp(-2)$ or $\exp(-0.4)$
c_{qpo}	...	$\exp(1) \text{ s}^{-1}$
f_{qpo}	...	20 Hz
A	3	3
t_0	0.5 s	0.5 s
σ	0.2 s	0.2 s

Note. a_{qpo} is listed for both the low- and high-amplitude simulations.

define the QPO Bayes factor as

$$\text{BF}_{\text{qpo}} = \frac{Z(d|k_{\text{qpo+rn}}, \mu)}{Z(d|k_{\text{rn}}, \mu)}, \quad (19)$$

where the numerator and denominator are the respective evidences given either kernel function and the same mean model μ . We expect that lnBF_{qpo} should generally be positive if we created the data using $k_{\text{qpo+rn}}$ and negative otherwise. We create the data on a 1 s interval sampled equidistantly at 256 Hz using the parameters listed in Table 3. We choose equidistant sampling in this case because prior choices do matter for model selection and we want to validate model selection with the prior ranges we define in Section 2.6 and apply on real data. Specifically, the Nyquist frequency for unevenly sampled data can be orders of magnitude greater than for evenly sampled data, thus greatly enlarging the prior volume of f_{qpo} .

In Figure 5 we show the result of performing this analysis for 1000 simulated light curves produced for each set of parameters. The Bayes factor generally shows a relatively large scatter, in line with previous explorations of the behavior

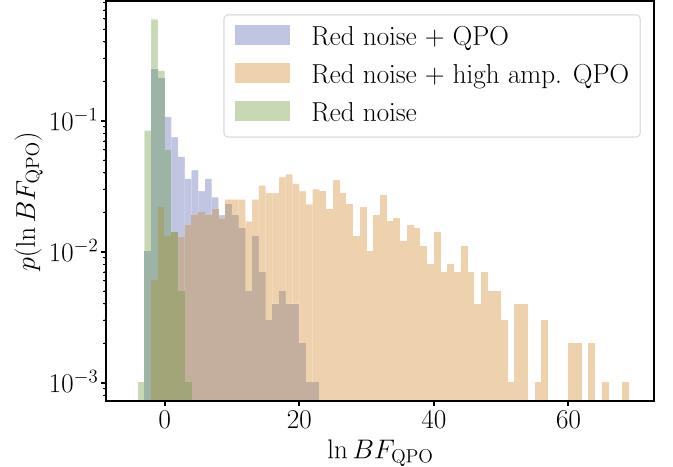


Figure 5. Normalized histogram of the lnBF_{qpo} values obtained from data sets produced with identical $k_{\text{qpo+rn}}$ (blue), $k_{\text{qpo+rn}}$ with a higher-amplitude QPO (orange), and k_{rn} (green). A total of 52.9% and 97.2% of the lower- and higher-amplitude QPOs yield $\text{lnBF}_{\text{qpo}} > 0$ (the logarithmic scale is somewhat deceiving). This shows that there is substantial spread in terms of the possible Bayes factors.

of model comparison via Bayes factors with noisy data (Jenkins 2014; Joachimi et al. 2021). The light curves containing a low-amplitude QPO have a positive lnBF_{qpo} 56.7% of the time, but some light curves yield $\text{lnBF}_{\text{qpo}} > 20$, indicating very high significance. High-amplitude QPOs, on the other hand, yield a positive lnBF_{qpo} 97.2% of the time, which indicates that we can almost always correctly identify QPOs with a sufficient amplitude. Even for high-amplitude QPOs the Bayesian analysis sometimes favors k_{rn} owing to its smaller prior volume. The scatter in the Bayes factor for the large-amplitude QPO is, perhaps counterintuitively, larger than for the small-amplitude case. This is in line with results by Joachimi et al. (2021), who suggest that the larger variance in Bayes factors is driven by an increase in the scatter in the evidence, as well as an expected decorrelation of evidences for the two models under comparison. Figure 6 demonstrates why there is such a high variance in the lnBF_{qpo} values we obtain. Despite being drawn from the same multivariate Gaussian distribution, QPOs can have vastly different amplitudes in the actual light curves, and even high-amplitude QPOs can be missed.

On the other hand, 94% of the data sets containing only red noise yield a $\text{lnBF}_{\text{qpo}} < 0$, though the distribution is much narrower. Only one simulated red-noise data set achieved a $\text{lnBF}_{\text{qpo}} > 2$. This narrower distribution is likely because k_{rn} is a limiting case of $k_{\text{qpo+rn}}$. A data set created using k_{rn} thus always fits well with $k_{\text{qpo+rn}}$, and the k_{rn} kernel model achieves a preference due to its smaller prior volume. If there is significant oscillatory behavior, k_{rn} cannot provide a good fit. In practical applications, and in particular in large sample studies, it may be advisable to not rely on standard prescriptions for significance thresholds (e.g., Kass & Raftery 1995), but rather calibrate false-alarm probabilities via simulations of a realistic model. Full evidence calculations for large samples of simulations may be prohibitively expensive, but approximate prescriptions based on a maximum likelihood analysis exist (Joachimi et al. 2021).

Testing the impact of modifying c_{qpo} and f_{qpo} on the expected significance in the form of a lnBF_{qpo} is expensive to perform empirically, so we will focus on a qualitative

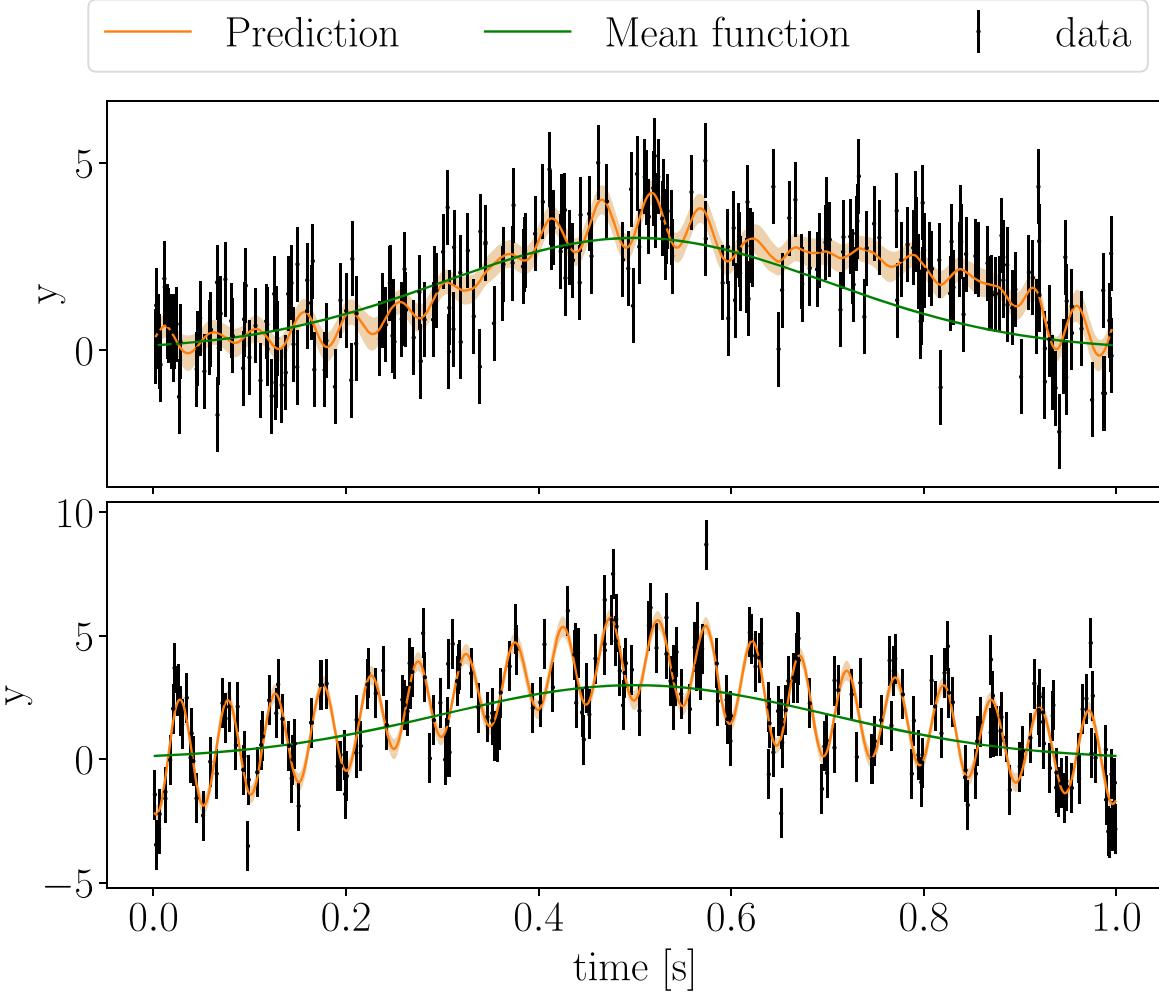


Figure 6. The least (top) and most (bottom) significant QPO in the higher QPO amplitude sets in the model selection study in Section 3.2. Despite having the same underlying parameters, the QPO is visibly much more pronounced in the bottom panel.

discussion. Generally, a smaller c_{qpo} corresponds to a higher-quality QPO, which should be more easily identifiable and thus yield a higher lnBF_{qpo} . Higher-frequency QPOs should also be more easily identifiable. At higher frequencies, the QPO has more total periods within a fixed length light curve. Additionally, if c_{qpo} is fixed, the QPO will decay by less within a single period and thus have a higher oscillation quality. This also highlights that identifying low-frequency QPOs is difficult because there are too few recorded oscillation periods within the data.

Overall, the results indicate that if we find a $\text{lnBF}_{\text{qpo}} > 2$, it is very unlikely that we have seen a false positive. We are also far more likely to miss real signals with weak QPOs than find false-positive QPOs. This tilt is preferable if we apply the methods on data where a false positive would be particularly detrimental for our understanding of astrophysics, for example, if we consider astrophysical sources for which we do not yet know if they may contain QPOs or not.

3.3. Nonstationary Time Series

Huebner et al. (2021) have shown that inappropriately applying periodogram-based methods on astrophysical time series can vastly overestimate their significance. Specifically, periodograms only yield statistically independent frequency bins if the time series is stationary and homoscedastic. Bias

occurs because bins close to the QPO frequency are not pairwise statistically independent. Thus, any analyses, including Bayesian inference methods based on the Whittle likelihood, that combine the power from multiple bins that make up the QPO is bound to overestimate the QPO's significance.

Since the Whittle likelihood directly derives from the assumption that the time series is Gaussian, it is intuitive that the nonstationarity bias should also exist in GP likelihoods. GPs account for heteroscedasticity and deterministic trends, but the stationary kernels we are using cannot account for nonstationary behavior in the noise as described in Huebner et al. (2021) with periodograms. We laid out in Section 2.4 how we can create a simple, nonstationary GP model using CELERITE.

Huebner et al. (2021) provide an extensive set of simulations to demonstrate the nonstationarity bias in simulated light curves. We can use one of their light curves and results to perform a comparative study on how we can address the nonstationarity bias with GPs. Specifically, Huebner et al. (2021) created the simulated light curve in Section 4.4 by producing red noise and a 1 Hz QPO from -10 to 10 s using the Timmer & König (1995) method. Next, they applied a Hann window to ensure a smooth turn-on from zero and added Gaussian white noise between -200 and 200 s to create a nonstationary time series. Huebner et al. (2021) showed that

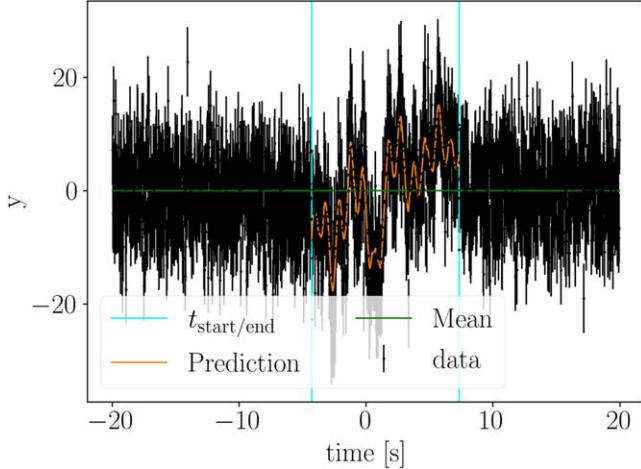


Figure 7. Maximum likelihood fit of the data (black) from Section 4.4 in Huebner et al. (2021) with a nonstationary GP model. We use a constant zero mean model (green). The teal lines indicate the maximum likelihood $t_{\text{start}/\text{end}}$ of the red noise and QPO GP. We underestimate the real duration of the GP, which is 20 s. This is because Huebner et al. (2021) used a Hann window in the data creation, which suppresses the amplitude closer to ± 10 s.

carrying out periodogram-based analysis on segments longer than the window containing the signal yields far higher $\ln BF_{\text{qpo}}$ owing to the nonstationarity bias.

We apply the nonstationary version of the GP models from Section 2.4 and the regular stationary version on simulated data from Section 4.4 in Huebner et al. (2021). Thus, we can quantify the bias that occurs as a result of the nonstationary change in noise at ± 10 s, as a function of the length of the white-noise segments added on either side of the segment with the signal, parameterized as an extension factor x .

Figure 7 shows the maximum likelihood fit of the time series from -20 to 20 s using the nonstationary GP modification we introduce in Section 2.4. The $t_{\text{start}/\text{end}}$ parameters delineate where the transition between GP and white noise occurs. Because of this, the nonstationary GP model yields a near-constant $\ln BF_{\text{qpo}}$ regardless of how much the time series is extended by white noise, as we show in Figure 8. This result for the nonstationary GP is desirable, since adding white noise on either end of a time series should not alter the information content about the QPO in the time series. On the other hand, the $\ln BF_{\text{qpo}}$ values that we obtain for stationary GP models or the periodogram rise first owing to nonstationarity bias and then fall off again as the QPO is covered by increasing white noise. Empirically, we find that the bias is much stronger for periodograms. This difference is likely because the kernel functions in the GP models effectively restrict the bias to a local timescale of $1/c_{\text{qpo}}$, whereas the periodogram uses a Fourier transform that calculates the power in each frequency bin from the entire time series. Thus, far away from where the QPO occurs, no additional bias can accrue for the GP because we compare model fits of k_{rn} and $k_{\text{qpo+rn}}$ on white-noise data. We show in Figure 9 that the nonstationary $k_{\text{qpo+rn}}$ GP is indeed preferred over the stationary one except for the segment between ± 10 s.

As we show in Figure 10 for the segment spanning between ± 20 s, the two-dimensional $t_{\text{start}/\text{end}}$ posterior is well constrained. We note that these parameters have multimodal posterior distributions that are hard to sample. The posterior we infer for t_{start} is also not consistent with the true start time of -10 s in Huebner et al. (2021). This inconsistency is likely

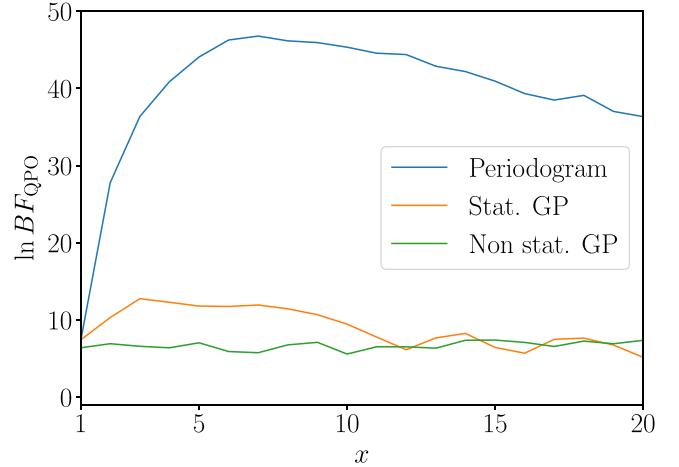


Figure 8. $\ln BF_{\text{qpo}}$ for a segment extending from $-10x$ to $10x$ s for a variable extension factor x . The signal is only contained in the central 20 s, such that for $x = 1$ the light curve contains only signal, for $x = 2$ the light curve contains 10 s of white noise both before and after the segment with the signal, and so on. We compare three QPO detection methods: a standard periodogram-based method (blue), a stationary GP model (orange), and the nonstationary extension to the GP model (green). While the periodogram and the stationary GP model are affected by the nonstationarity bias, the nonstationary GP model yields a near-constant $\ln BF$.

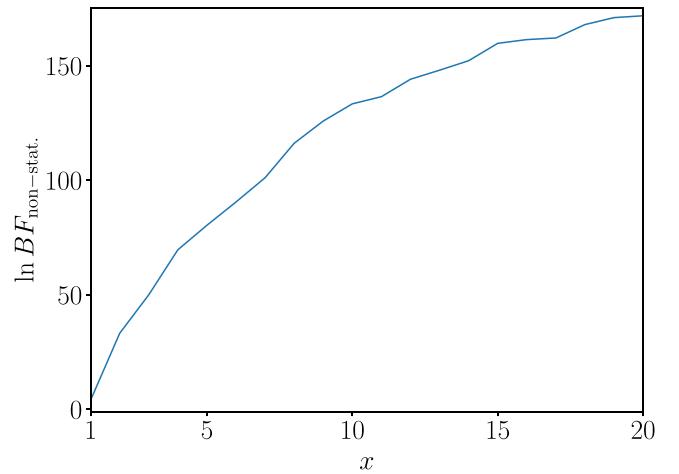


Figure 9. $\ln BF_{\text{nonstat}}$ of the nonstationary $k_{\text{qpo+rn}}$ GP model relative to the stationary GP for a segment between $-10x$ and $10x$ s. The nonstationary model is preferred since it correctly models the change from red noise, white noise, and QPO to just white noise.

because of how Huebner et al. (2021) created the data. Close to ± 10 s the Hann window suppresses red noise and QPO.

4. Solar Flares: Hares and Hounds

Quasi-periodic pulsations (QPPs)—as QPOs are called in the solar flare literature—have been a regularly reported feature of solar flares for decades (see Van Doorsselaere et al. 2016; Zimovets et al. 2021 for recent reviews). They have also been observed, albeit less often, in stellar flare light curves. They can occur in a wide range of wavelengths and hence in both thermal and nonthermal flare emission, from radio waves (e.g., Grechnev et al. 2003; Melnikov et al. 2005; Inglis et al. 2008; Kupriyanova et al. 2016) to ultraviolet emission (e.g., Brosius et al. 2016), X-rays (Parks & Winckler 1969; Kane et al. 1983; Asai et al. 2001; Hayes et al. 2016, 2019), and even gamma-rays (Nakariakov et al. 2010).

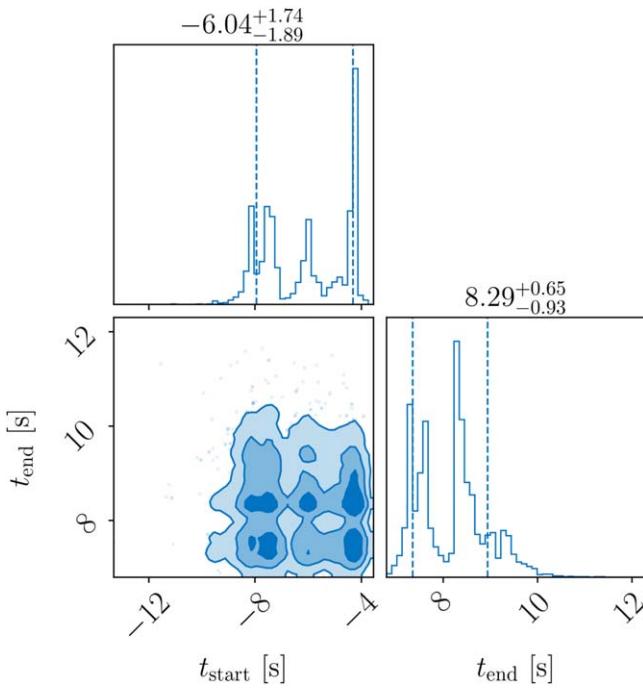


Figure 10. 2D $t_{\text{start}}/t_{\text{end}}$ posterior for $x = 2$ for the data from Section 4.4 in Huebner et al. (2021), which we show in Figure 7. These parameters are hard to sample compared to the kernel and mean model parameters, as they tend to be multimodal.

On the Sun, the typical oscillation periods observed range from a few seconds up to several minutes. They can occur at any time during a flare, from the impulsive phase of energy release to the gradual decay phase. QPPs are particularly interesting observations in flares because they provide a potential diagnostic of fundamental flare energy release processes and the flaring plasma properties. The most established explanations for solar flare QPPs include periodic or bursty magnetic reconnection and modulation of solar flare plasma via magnetohydrodynamic waves. A full discussion of potential emission mechanisms may be found in McLaughlin et al. (2018) and Zimovets et al. (2021).

Historically, many different methods have been used to detect QPPs in solar flares. Recently, the robustness of different analysis methods has been studied by various authors (e.g., Auchère et al. 2016; Dominique et al. 2018; Broomhall et al. 2019). In particular, Broomhall et al. (2019) performed a comparative study between different QPP detection methods. In this study, different methods were implemented and deployed by experts, who were blinded to which light curves did contain QPOs. One of the methods used CELERITE GP models in combination with a polynomial flare mean model (Davenport et al. 2014) and a detrending method. The participants analyzed 100 light curves; 60 contained a QPP. The GP-based model in Broomhall et al. (2019) flagged 52 light curves for QPPs, only 29 of which actually contain a QPP, and thus performed worse than picking events at random. However, Broomhall et al. (2019) used the stochastically driven harmonic oscillator (SHO) kernel function as its sole model and implemented a heuristic detection criterion requiring that the oscillation period of the maximum likelihood estimate is between 3 and 200 units of time (data points are separated by 1 unit of time with 300 data points per flare). Without additionally including information about the quality factor Q

(i.e., the width) of the QPO, it is unclear whether this condition alone could determine whether or not the SHO model is consistent with a QPO.

In the limit of low-quality factors, the SHO becomes aperiodic, which should be used as a criterion to reject the QPP hypothesis. Specifically, Foreman-Mackey et al. (2017) showed that the SHO has commonly used aperiodic forms in the limit of $Q \leq 1/2$ and $Q = 1/\sqrt{2}$. We conclude that the differences in performance between those in Broomhall et al. (2019) and our approach presented above are likely rooted in a combination of the different model assumption (e.g., the detrending step) and the differences in how the significance is assessed.

We also note that the Hares and Hounds data diverge significantly from our assumptions about the data. First, Broomhall et al. (2019) created data with an unknown white-noise component, so we add a white-noise kernel term. Second, the light curves are created with several different mean models. Specifically, each simulation has flares based on Gaussian or exponential components, which are repeated once or twice in some cases, and the data have an offset and a linear trend. Most importantly, while the QPP is intermittent and only appears on the tail side of the flare, red noise within these light curves persists for the entire segment. The data are thus nonstationary in a way that we cannot perfectly model using the modifications to CELERITE we introduce in Section 2.4.

Broomhall et al. (2019) use an exponentially damped sinusoid as a QPP model. Using an analytical function instead of a random process underlying the QPP also has implications for the results from the frequency-domain-based methods in Broomhall et al. (2019). Specifically, frequency bins around the QPP frequency are not statistically independent.

Performing an analysis on all light curves can provide a point of comparison between our method and the different methods laid out in Broomhall et al. (2019). However, details such as mean models and noise properties were unknown to the participants in their study.

For our analysis, we restrict ourselves to a single skewed exponential flare mean model plus an offset and reanalyze the 100 light curves from the second Hares and Hounds round. We again compare the k_{rn} and $k_{\text{qpo+rn}}$ hypotheses using Bayesian inference, but we also add a white-noise jitter term to the kernel to account for the unknown noise. We also opt for removing the time series left of the maximum since the QPPs are explicitly only present in the decay phase. Some study participants also made the same decision. We display one of the fitted time series that contains a QPP in Figure 11.

The data set contains 40 light curves with a QPP and 60 without a QPP. Additionally, there are some light curves with repeating flare shapes, and the flare shapes can be either Gaussian or exponential.

We find four light curves with $\ln BF_{\text{qpo}} \geq 2$, all of which contain a QPP. Given our findings in Section 3.2, in which only 1 out of 1000 red-noise signals caused a $\ln BF_{\text{qpo}} > 2$, this is a reasonable and conservative detection threshold. The most significant QPP has $\ln BF_{\text{qpo}} = 9.7$. There are three more light curves with $0 < \ln BF_{\text{qpo}} < 2$, none containing a QPP. We thus find fewer QPPs than other models Broomhall et al. (2019) explored, specifically AFINO (Inglis et al. 2015, 2016), which found eight QPPs without a false detection. However, the model setup here is a significant improvement over the implementation of the GP model in Broomhall et al. (2019),

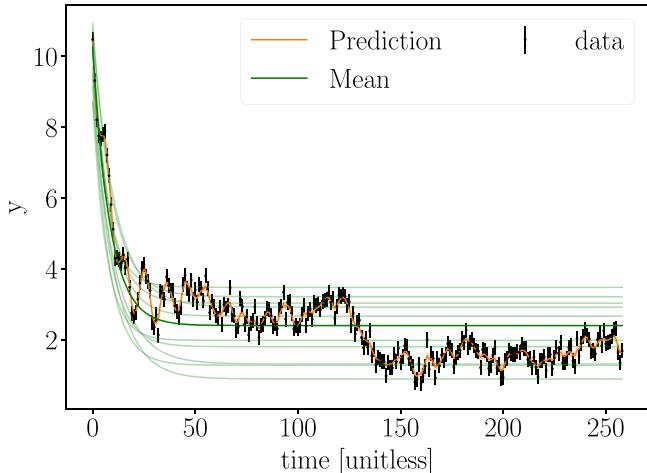


Figure 11. Maximum likelihood fit of the Hares and Hounds data set 404267 using a skewed exponential and the $k_{\text{qpo}+\text{rn}}$ kernel. The data contain a QPP, of which a few cycles can be seen by eye for $t < 50$. We show the mean function from the maximum likelihood sample (dark green) and 10 other samples from the posterior (light green). The orange curve is the prediction based on the maximum likelihood sample and the 1σ confidence band.

suggesting that further development of our model taking into account the specific idiosyncrasies of solar flare data will provide a powerful tool for QPP searches in these light curves.

5. Real Data

5.1. Gamma-Ray Bursts

There is ample speculation about the possibility for QPOs in long GRBs (Masada et al. 2007; Ziaeepour & Gardner 2011), and there have been some recent claims about possible detections (Tarnopolski & Marchenko 2021). One of the most tantalizing events is GRB090709A, a long GRB for which Cenko et al. (2010) found a marginal 2σ 8.06 s period QPO in the prompt emission. Independent analyses have since found similar significance levels between 2σ and 3.5σ (Iwakiri et al. 2010; Ziaeepour & Gardner 2011; Dichiara et al. 2013; Guidorzi et al. 2016). GRB090709A had its strongest emission for about 100 s and a visible afterglow for several hundreds of seconds afterward. Cenko et al. (2010) primarily relied on the Swift light curve for their analysis, though they also considered the Suzaku light curve, which did not increase the significance. Further analyses of the Swift and XMM-Newton light curves by De Luca et al. (2010), who interpreted GRB090709A as a distant, standard, long GRB, also showed no periodicity above the 3σ level. Konus and SPI-ACS also recorded light curves of GRB090709A, though it is not firmly established whether the combined data of all instruments would yield higher significance (Iwakiri et al. 2010). Cenko et al. (2010) and De Luca et al. (2010) detrend the light curve before calculating the periodogram to enhance the periodicity.

For our reanalysis of this event, we focus on the Swift light curve, though joint inference of parameters across multiple observations is, in principle, possible. We select 107 s of the overall light curve as suggested by Cenko et al. (2010). We carry out inference with the models we define in Section 2.3 and allow a constant positive offset with all these models. Furthermore, we analyze the data with one to three flare components to see whether there is a substantial improvement when using a mean function with multiple components. As

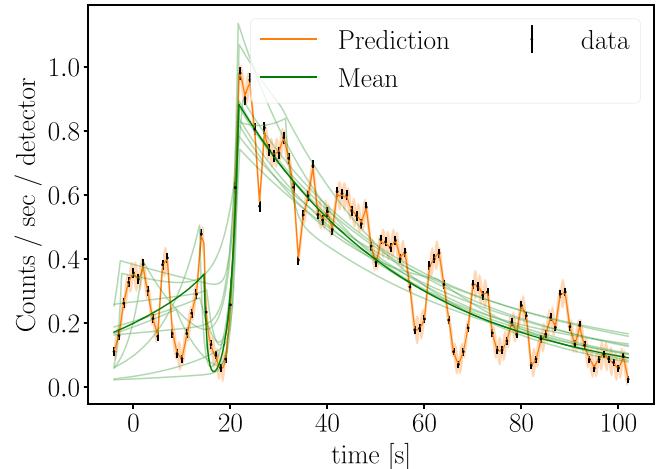


Figure 12. GRB090709A fit using two skewed exponentials and the $k_{\text{qpo}+\text{rn}}$ kernel. We show the mean function from the maximum likelihood sample (dark green) and 10 other samples from the posterior (light green). The orange curve is the prediction based on the maximum likelihood sample and the 1σ confidence band.

with our framework described in Section 2.2, we carry out Bayesian model selection between k_{rn} and $k_{\text{qpo}+\text{rn}}$.

We perform Bayesian inference using DYNESTY with 1000 live points and the random walk sampling method. Figure 12 shows the data and a fit of the skewed exponential mean function with two flares, as well as the prediction curve generated by CELERITE using the maximum likelihood posterior sample. In the top panel of Figure 13, we display the obtained BF_{qpo} for various mean models. We generally find that adding more components to our mean models does not meaningfully change the BF_{qpo} and only somewhat increases the BF_{qpo} relative to a constant zero mean model (bottom panel of Figure 13). This hints that already a single-component model is a reasonable model for the overall trends in the data. Furthermore, there is some difference between the evidences between the individual mean models. The difference is relatively minor between the skewed exponential and Gaussian model, but the FRED model yields substantially lower evidences. Overall, we find that the lnBF_{qpo} values fall between -1.5 and 1.1 . We find that we obtain a well-constrained $P = 1/f = 8.31_{-0.33}^{+0.42}$ s (68% CI) posterior (see Figure 14) that is consistent with the previously reported $P = 8.1$ s period (Cenko et al. 2010; De Luca et al. 2010). The low BF_{qpo} indicate that we should not consider this to be a confident detection, though in the future combining a significance criterion based on the Bayes factor with an *effect size*, for example, based on how well constrained the posterior for the QPO frequency is, could improve our ability to detect QPOs in cases where the Bayes factor alone is inconclusive.

5.2. Magnetar Bursts

Magnetars can show strong bursting behavior, ranging from a series of low-energetic recurrent bursts to rare giant flares, of which we have observed few so far. Some mechanisms may trigger a QPO in a magnetar burst, such as torsional Alfvén oscillations (Levin 2007; Sotani et al. 2008), but the observational evidence remains scant. So far, there have only been definitive detections in the SGR 1806–20 and SGR 1900+14 giant flares (Israel et al. 2005; Strohmayer & Watts 2005, 2006;

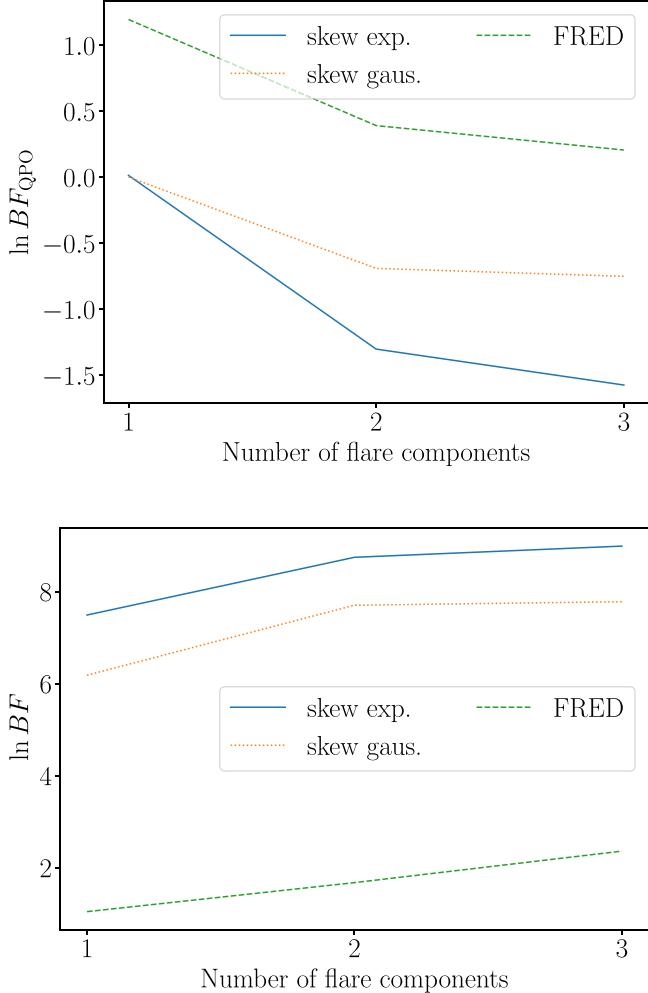


Figure 13. Results for GRB090709A. Top: $\ln BF_{\text{QPO}}$ for different mean models and different number of flare components. We find in all instances that neither k_{rn} nor $k_{\text{qpo+rn}}$ is strongly preferred for any specific configuration. Bottom: $\ln BF$ for different mean models using the $k_{\text{qpo+rn}}$ model relative to a constant zero mean model.

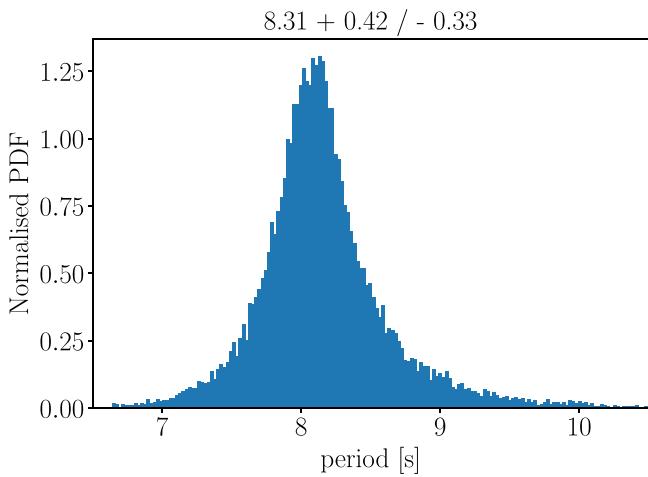


Figure 14. Period posterior for GRB090709A using two skewed exponentials. The period is well constrained, $P = 1/f = 8.31_{-0.33}^{+0.42}$ s (68% CI), and consistent with the results from Cenko et al. (2010) and De Luca et al. (2010), but it has wide tails with low probability.

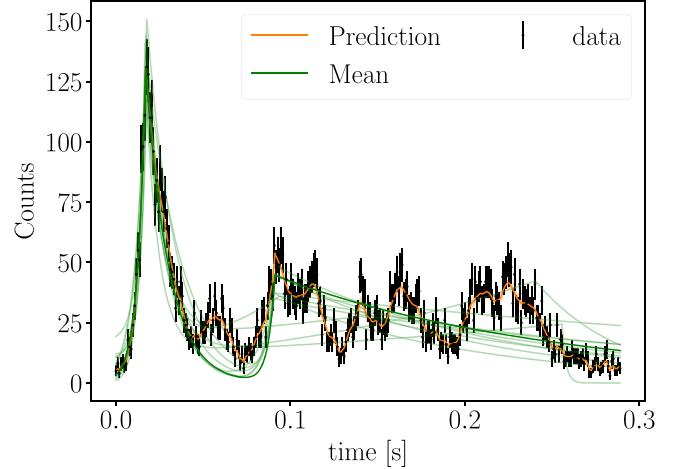


Figure 15. Magnetar burst 080823478 fit using two skewed exponentials and the $k_{\text{qpo+rn}}$ kernel. We show the mean function from the maximum likelihood sample (dark green) and 10 other samples from the posterior (light green). The orange curve is the prediction based on the maximum likelihood sample and the 1σ confidence band.

Watts & Strohmayer 2006) and in some recurring smaller bursts (Huppenkothen et al. 2012, 2014a, 2014c).

We demonstrate our method on a single magnetar burst 080823478 observed with the Fermi Gamma-ray Burst monitor from SGR 0501+4516, a magnetar that was first discovered in 2008 by the Swift satellite (Rea et al. 2009). For details on the data processing, see Huppenkothen et al. (2012). As we can see in Figure 15, the burst has a shape that, by eye, appears to be periodic, with three roughly equally spaced peaks after the main peak. Following the same steps we have used for the GRB, we show in Figure 16 that the $\ln BF_{\text{QPO}}$ values vary between -0.4 and 0.9 depending on the mean model and there is thus no strong evidence for the presence of a QPO. Moreover, the period posterior in Figure 17 shows that we are also unable to consistently constrain f_{qpo} . This finding is in agreement with the conclusions of Huppenkothen et al. (2012), which used this burst as a template for studies of red noise in magnetar bursts and also found no periodicity.

Alternatively, we also test a constant mean model, meaning that we model all variability as arising owing to a stationary stochastic process. In that case, $k_{\text{qpo+rn}}$ is strongly preferred over k_{rn} ($\ln BF_{\text{QPO}} = 6.8$); however, this mean model is strongly disfavored compared to all the other mean models we have tested ($\ln BF = -21$ compared with the single skew exponential model). These results highlight the benefit of using mean models, as they help us disentangle whether variability arises from the overall shape of the burst or owing to a stochastic process.

5.3. Giant Magnetar Flare

Multiple X-ray observatories, including the Rossi X-ray Timing Explorer (RXTE; Hurley et al. 2005; Palmer et al. 2005), recorded the SGR 1806-20 giant flare on 2004 December 27. The flare, which lasted about 380 s, has been extensively studied for the presence of QPOs (Israel et al. 2005; Strohmayer & Watts 2005, 2006; Watts & Strohmayer 2006; Huppenkothen et al. 2014b; Miller et al. 2019). Notably, QPOs were associated with specific phases in the 7.56 s rotational period (Strohmayer & Watts 2006), which indicate that the flare

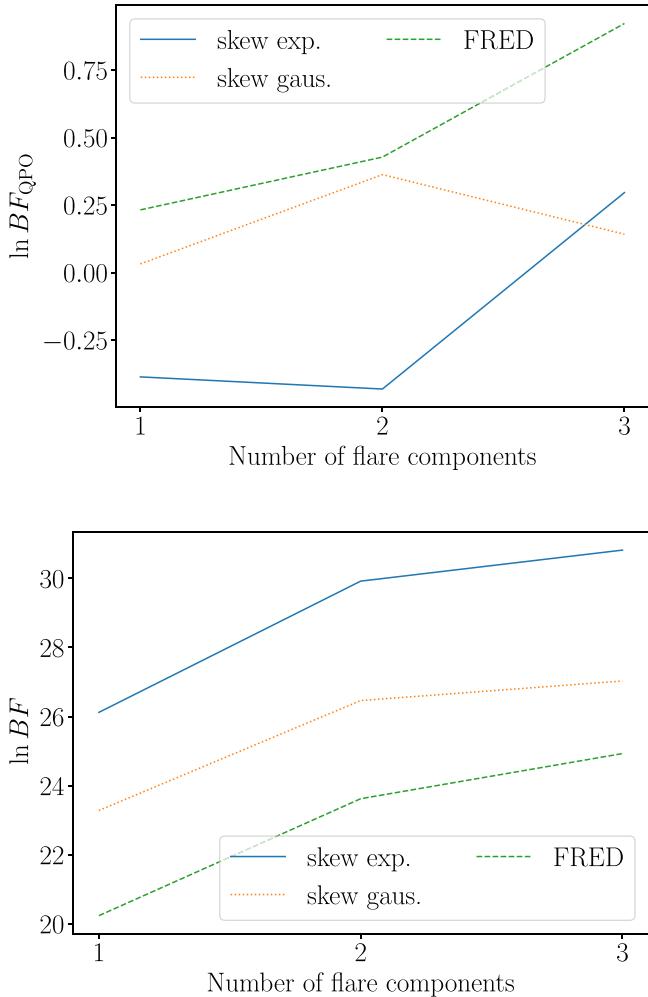


Figure 16. Results for the 080823478 burst. Top: $\ln BF_{\text{QPO}}$ for different mean models and different number of flare components. We find in all instances that neither k_m nor $k_{\text{qpo}+m}$ is strongly preferred for any specific configuration. Bottom: $\ln BF$ for different mean models using the $k_{\text{qpo}+m}$ model relative to a constant zero mean model.

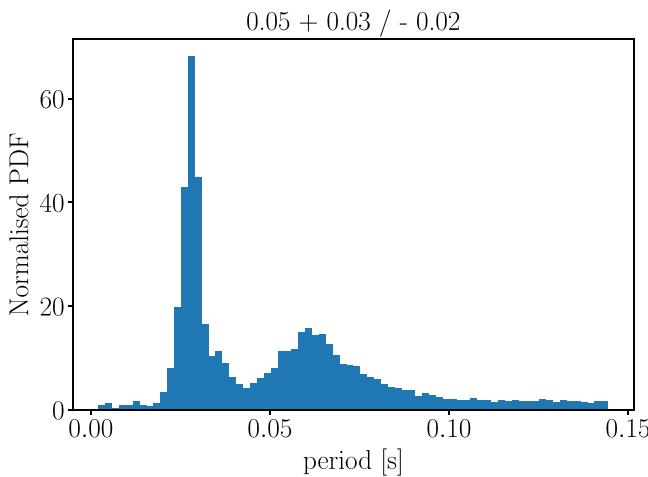


Figure 17. Period posterior for the 080823478 burst. The posterior is not constrained and has support across the entire prior space. This result is not representative of all runs we performed. Sometimes the period is more narrowly constrained, but $k_{\text{qpo}+m}$ is still disfavored.

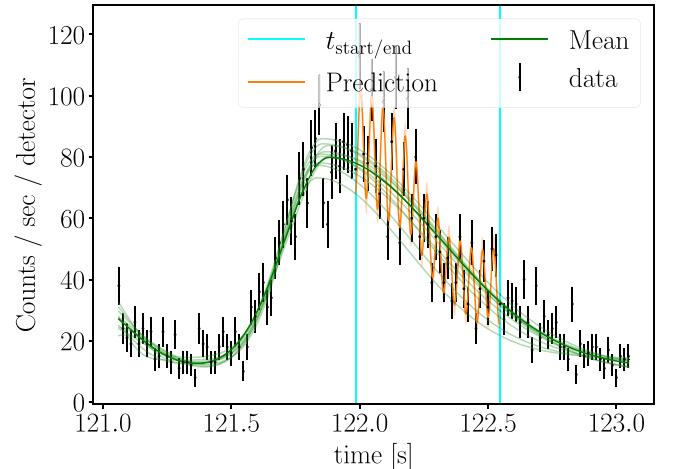


Figure 18. Giant flare maximum likelihood fit from our selected 2 s segment using two skewed Gaussians using the nonstationary $k_{\text{qpo}+m}$ model. We show the mean function from the maximum likelihood sample (dark green) and 10 other samples from the posterior (light green). The orange curve is the prediction based on the maximum likelihood sample and the 1σ confidence band.

is associated with a specific region on the magnetar surface that is turning in and out of view. Strohmayer & Watts (2006) specifically identified the most significant QPOs at 18, 26, 29, 93, 150, 625, and 1837 Hz. Initial attempts used averaged periodograms created by selecting the same time interval in subsequent rotational phases of the magnetar. However, the nonstationary nature of the light curve within these time intervals may bias these results.

Miller et al. (2019) took a more systematic approach in analyzing the giant flare by using a 1 s sliding window that they moved across the flare light curve in 0.945 s steps. This way, they were able to locate when specific QPOs were occurring, though the time resolution is limited to ≈ 1 s.

Given the findings of the nonstationarity bias in Huebner et al. (2021), identifying when QPOs start and end in the flare light curve is an essential key to assessing their significance. We can use our nonstationary model from Section 2.4 to make quantitative statements for individual QPOs.

For this study, we focus on a specific 23 Hz QPO that was located in the RXTE light curve by Miller et al. (2019) to occur about 122.060 s after the beginning of the light curve at 2004 December 27, 21:30:31.378 UTC. To see whether we can localize when the QPO occurs, we use the nonstationary model and select a 2 s segment starting at 121.060 s. Miller et al. (2019) also reported a 92 Hz QPO to occur at the same segment. Thus, we choose a relatively coarse binning of the time-tagged events of 64 Hz so that we can focus on the lower-frequency QPO. As we display in Figure 18, we can constrain the time when the QPO occurs within the light curve. The inferred QPO mean frequency from the nonstationary model is $f_{\text{qpo}} = 22.90^{+0.45}_{-0.69}$ Hz (68% CI). We find that the QPO lasts for $0.81^{+0.23}_{-0.21}$ s (68% CI) and is primarily located on the top and tail side of the peak, though the posterior as shown in Figure 19 is multimodal, with the maximum being closer to 0.5 s. Comparing models with ($k_{\text{qpo}+m}$) and without (k_m) a QPO, we obtain $\ln BF_{\text{QPO}} = 6.8$ using the stationary model. Including our prescription for nonstationarity and again comparing a model

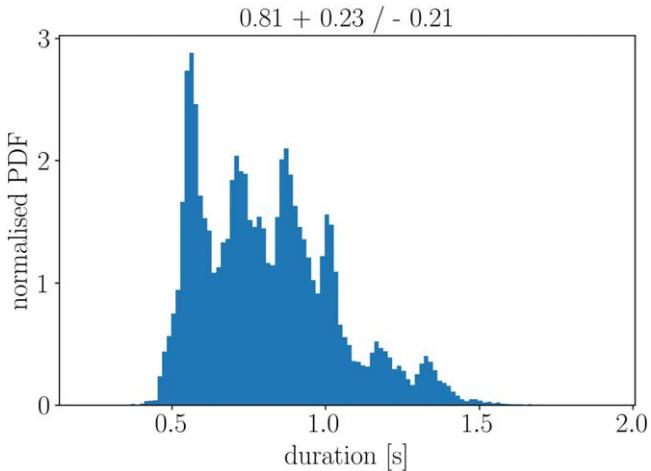


Figure 19. Posterior distribution of QPO duration ($t_{\text{end}} - t_{\text{start}}$) for the fit in Figure 18.

including a QPO to one without, we obtain $\ln \text{BF}_{\text{QPO}} = 8.8$. Both yield strong evidence for the presence of a QPO. Finally, we can also compare the stationary and nonstationary models where both contain a QPO (using the $k_{\text{QPO+rn}}$ kernel), and we find that the nonstationary model is preferred over the stationary one with $\ln \text{BF} = 5.1$.

Expanding on these results by analyzing more segments may allow us to explore the temporal structure of the QPOs in more detail. The specific segment we have considered contains one of the most significant QPOs in the entire flare. More marginal QPOs are harder to find and characterize. Eventually, hierarchical models may allow us to better understand the nature of QPOs in giant flares by taking a broader view of the entire 400 s of data. Concretely, we may constrain where QPOs in the pulse period start and end and use information from multiple segments to find all frequency modes.

6. Discussion and Outlook

We introduce a new method to search for and analyze QPOs in the presence of red noise using a combination of GP modeling and Bayesian inference with evidence calculation. Our work was motivated by the large interest in searching for and characterizing quasi-periodicities in fast transients, in particular in the X-ray light curves of solar flares, magnetar bursts, and GRBs. All three share characteristics that make QPO searches challenging: fast, nonstationary variability, the requirement to perform searches down to low frequencies, and the lack of a physically motivated generative model. However, the method is designed to be generally applicable (within the limits laid out in Section 2.5): mean functions and covariance function admit drop-in replacements depending on the application.

We can model all aspects of a transient in the time domain using this method. We use a set of phenomenological mean models to describe the overall flare shape and different GP models to model either red noise or a combination of red noise and QPOs. Using studies on simulated data, we show that we can accurately estimate the true parameters, model simple nonstationary GPs, and avoid some of the biases that nonstationary behavior causes when applying Fourier-based methods. We also show that we can distinguish QPOs from red noise, assuming that the QPOs have a sufficient amplitude to be

detected. We demonstrate that one can easily apply this framework with few modifications on many different astrophysical transients. Overall, the results from astrophysical data agree with the previously reported results in the literature. We find that the application of GP methods in the analysis of X-ray time series is thus very promising, and specifically, we found that the method helps to constrain when QPOs occur within the giant flare light curve of SGR 1806–20.

While this paper provides an overview of how GPs can be applied in QPO searches in transient light curves, future work will include a number of improvements and extensions, in particular to make the method more powerful for specific sources and applications. CELERITE limits us to a narrow class of kernel functions, which are mixtures of stochastically driven, damped harmonic oscillators, motivated by the typical variability seen in the light curves of stars (Foreman-Mackey et al. 2017). Mixtures of sufficient components are capable of representing a wide range of underlying power spectra, but at the loss of physical interpretability for a given problem. However, our work here was in part motivated by the lack of physical models capable of describing the light curves of many types of high-energy transients. Attempts have been made at a general characterization of the power spectrum of GRBs with both Fourier-based methods (e.g., Beloborodov et al. 2000) and wavelets (Tarnopolski & Marchenko 2021), which suggest that GRBs can generally have a wide range of power-law indices and are sometimes better characterized by broken power laws. However, these works often use methods that implicitly assume stationarity, and thus it is unclear how well these results translate into informative priors on the choice of covariance function. While here we primarily consider the GP model as a method for searching for QPOs, we note that it can also help us understand the broader variability properties in transients. Future work might explore different models for both mean model and covariance function across a large sample of transients of a given source class, using a similar framework based on Bayes factors to what we have used here to compare these different model choices. Aside from providing new insights into the source physics, this work would also provide a better baseline model to use for QPO searches.

There have been recent claims of detections of QPOs in FRBs observed with the CHIME instrument (Andersen et al. 2021; Pastor-Marazuela et al. 2022). Pastor-Marazuela et al. (2022) specifically consider FRB 20201020A, for which they find a QPO at the 2.5σ level. FRB studies currently use a mix of periodogram-based methods and more bespoke methods for which a systematic application of GPs would provide a valuable cross-check. FRBs are observed in the radio, where instrument noise may behave more similarly to the Gaussian assumption inherent in the GP framework than photon-counting data do. In addition, FRBs are similar to other high-energy transients in their lack of generative, physically motivated models of source emission and fast, nonstationary variability, making them another interesting future application for the methods we have laid out in this paper.

The framework introduced here can be extended by using a more general GP framework with fast solvers. For example, the HODLR solver, of which the GEORGE package contains an implementation, operates in $\mathcal{O}(N \log^2 N)$ and thus can be run in acceptable time for the astrophysical data sets we have analyzed (Ambikasaran et al. 2016). Specifically, this would allow us to implement kernels such as the Matérn class of

covariance functions or the squared exponential covariance function, which allows us to model a larger class of noise processes. Wilkins (2019) explored the applicability of different standard prescriptions for covariance functions with the observed variability of AGN light curves. A similar approach—taking into account the nonstationary nature of transients—might be a valuable next step in exploring transient variability.

In Section 2.5, we lay out assumptions underpinning and further limitations of GPs. Specifically, we highlight the issues when analyzing Poissonian data with low photon count rates, which have to be addressed using modified methods (Adams et al. 2009; Flaxman et al. 2015). Extensions into nonstationary covariance functions would also be an interesting avenue for future work.

In many astrophysical sources, time variability is intricately linked to changes in the spectral energy distribution. In the study of X-ray binaries and AGNs, this has led to the development of Fourier-based *spectral-timing* methods aimed at understanding astrophysical processes through this inter-linked behavior (for an example of a connection between QPO phase and spectral properties in black hole X-ray binaries, see Ingram et al. 2016; for a review, see Ingram & Motta 2019). To our knowledge, they have not been applied in the study of high-energy transients, where QPO searches are usually performed across the entire relevant spectral range. On the other hand, QPO detection and characterization methods, including one based on GPs, do exist for stationary time series. In the future, combining these approaches and enabling QPO searches that account for a potential spectral variability in the presence or characteristics of the QPO may yield better detection power.

The transients considered here are primarily high-energy fast transients observed with X-ray instruments. These transients are generally observed at a fine, evenly sampled cadence. Studies of transients observed over longer periods of time, such as GRB afterglows, kilonovae, and supernovae, may be able to make use of another, powerful property of the GP: its relative insensitivity to the sampling pattern. In these cases, realistic and physically motivated priors—especially on the QPO frequency—are of increased relevance, since priors based on the pseudo-Nyquist frequency may become unrealistically large.

The approach we have presented here has advantages over many standard QPO search methods in that it takes the transient’s nonstationarity into account and admits a range of parameterizations for both the mean function and covariance function. It is therefore flexible and applicable to a wide range of astrophysical transient phenomena. However, GPs are not the only possible approach for taking nonstationarity into account. Wavelets have been used in some studies to search for QPOs (e.g., Morris et al. 2010). Autoregressive moving-average (ARMA) models are popular in a wide range of scientific fields and are closely related to the covariance functions implemented in CELERITE. They can be adapted both to nonstationarity via autoregressive integrated moving-average (ARIMA; Box et al. 1970) models and to include periodic signals (periodic ARMA process; Adams & Goodwin 1995). A future study benchmarking the various approaches that include prescriptions for nonstationarity, including the method proposed here, (P)ARMA processes, and wavelets, would be very useful to the community.

We thank the anonymous reviewer for their valuable insights, which have substantially improved the manuscript. We also thank the AAS statistics editor for additional suggestions. This work is supported through ARC Centre of Excellence CE170100004. P.D.L. is supported through the Australian Research Council (ARC) Discovery Project DP220101610. D.H. is supported by the Women In Science Excel (WISE) program of the Netherlands Organisation for Scientific Research (NWO). This work was performed on the OzSTAR national facility at Swinburne University of Technology. The OzSTAR program receives funding in part from the Astronomy National Collaborative Research Infrastructure Strategy (NCRIS) allocation provided by the Australian Government.

Facilities: Fermi, Swift, RXTE.

Software: celerite (Foreman-Mackey et al. 2017), bilby (Ashton et al. 2019b; Romero-Shaw et al. 2020), dynesty (Speagle 2020), corner (Foreman-Mackey 2016).

ORCID iDs

Moritz Hübner <https://orcid.org/0000-0002-9642-3029>

Daniela Huppenkothen <https://orcid.org/0000-0002-1169-7486>

Paul D. Lasky <https://orcid.org/0000-0003-3763-1386>

Andrew R. Inglis <https://orcid.org/0000-0003-0656-2437>

David W. Hogg <https://orcid.org/0000-0003-2866-9403>

References

Adams, G., & Goodwin, G. 1995, *J. Time Ser. Anal.*, 16, 127

Adams, R. P., Murray, I., & MacKay, D. J. 2009, in ICML ‘09: Proc. of the 26th Annual Int. Conf. on Machine Learning (New York: ACM), 9

Ambikasaran, S., Foreman-Mackey, D., Greengard, L., Hogg, D. W., & O’Neil, M. 2016, *ITPAM*, 38, 252

Andersen, B. C., Bandura, K., Bhardwaj, M., et al. 2022, *Natur*, 607, 256

Angus, R., Morton, T., Aigrain, S., Foreman-Mackey, D., & Rajpaul, V. 2018, *MNRAS*, 474, 2094

Anscombe, F. J. 1948, *Biometrika*, 35, 246

Asai, A., Shimojo, M., Isobe, H., et al. 2001, *ApJL*, 562, L103

Ashton, G., Hübner, M., Lasky, P. D., et al. 2019a, *ApJS*, 241, 27

Ashton, G., Lasky, P. D., Gruber, V., & Palfreyman, J. 2019b, *NatAs*, 3, 1143

Auchère, F., Froment, C., Bocchialini, K., Buchlin, E., & Solomon, J. 2016, *ApJ*, 825, 110

Bar-Lev, S. K., & Enis, P. 1990, *Stat. Probab. Lett.*, 10, 95

Barret, D., & Vaughan, S. 2012, *ApJ*, 746, 131

Beloborodov, A. M., Stern, B. E., & Svensson, R. 2000, *ApJ*, 535, 158

Blei, D. M. 2014, *AnRSA*, 1, 203

Bond, J. C., Crittenden, R., Jaffe, A., & Knox, L. 1999, *CSE*, 1, 21

Bond, J. R., & Efstathiou, G. 1987, *MNRAS*, 226, 655

Box, G. E., Jenkins, G. M., Reinsel, G., et al. 1970, *Time Series Analysis* (New York: Wiley), 75

Broomhall, A. M., Davenport, J. R., Hayes, L. A., et al. 2019, *ApJS*, 244, 37

Brosius, J. W., Daw, A. N., & Inglis, A. R. 2016, *ApJ*, 830, 101

Castro-Tirado, A. J., Østgaard, N., Görgüç, E., et al. 2021, *Natur*, 600, 621

Cenko, S. B., Butler, N. R., Ofek, E. O., et al. 2010, *AJ*, 140, 224

Colaiuda, A., Beyer, H., & Kokkotas, K. D. 2009, *MNRAS*, 396, 1441

Colaiuda, A., & Kokkotas, K. D. 2011, *MNRAS*, 414, 3014

Cook, S. R., Gelman, A., & Rubin, D. B. 2006, *J. Comput. Graph. Stat.*, 15, 675

Covino, S., Landoni, M., Sandrinelli, A., & Treves, A. 2020, *ApJ*, 895, 122

Covino, S., Tobar, F., & Treves, A. 2022, *MNRAS*, 513, 2841

Czekala, I., Mandel, K. S., Andrews, S. M., et al. 2017, *ApJ*, 840, 49

Davenport, J. R., Hawley, S. L., Hebb, L., et al. 2014, *ApJ*, 797, 11

De Luca, A., Esposito, P., Israel, G. L., et al. 2010, *MNRAS*, 402, 1870

Delbridge, I. A., Bindel, D. S., & Wilson, A. G. 2020, in Proc. of the 37th International Conference on Machine Learning (New York: PMLR), 2453

D’Emilio, V., Green, R., & Raymond, V. 2021, *MNRAS*, 508, 2090

Dichiara, S., Guidorzi, C., Frontera, F., & Amati, L. 2013, *ApJ*, 777, 6

Dominique, M., Zhukov, A. N., Dolla, L., Inglis, A., & Lapenta, G. 2018, *SoPh*, 293, 61

Eyer, L., & Bartholdi, P. 1999, *A&AS*, **135**, 1

Flaxman, S., Wilson, A. G., Neill, D. B., Nickisch, H., & Smola, A. J. 2015, in 32nd Int. Conf. on Machine Learning (New York: PMLR), 607

Foreman-Mackey, D. 2016, *JOSS*, **1**, 24

Foreman-Mackey, D., Agol, E., Ambikasaran, S., & Angus, R. 2017, *AJ*, **154**, 220

Gabler, M., Cerdá-Durán, P., Font, J. A., Müller, E., & Stergioulas, N. 2011, *MNRAS*, **410**, L37

Gabler, M., Cerdá-Durán, P., Font, J. A., Müller, E., & Stergioulas, N. 2013, *MNRAS*, **430**, 1811

Gardner, J. R., Pleiss, G., Bindel, D., Weinberger, K. Q., & Wilson, A. G. 2018, in 32nd Conf. on Neural Information Processing Systems (Red Hook, NY: Curran Associates, Inc.)

Gelman, A., Carlin, J. B., Stern, H. S., et al. 2013, Bayesian Data Analysis (Boca Raton, FL: Chapman and Hall/CRC)

Gentle, J. E. 2009, Computational Statistics (Statistics and Computing) (New York, NY: Springer New York)

Gibson, N. P., Aigrain, S., Roberts, S., et al. 2012, *MNRAS*, **419**, 2683

Grechnev, V. V., White, S. M., & Kundu, M. R. 2003, *ApJ*, **588**, 1163

Grunblatt, S. K., Howard, A. W., & Haywood, R. D. 2015, *ApJ*, **808**, 127

Guidorzi, C., Dichiara, S., & Amati, L. 2016, *A&A*, **589**, A98

Hayes, L. A., Gallagher, P. T., Dennis, B. R., et al. 2016, *ApJL*, **827**, L30

Hayes, L. A., Gallagher, P. T., Dennis, B. R., et al. 2019, *ApJS*, **875**, 11

Hayes, L. A., Inglis, A. R., Christe, S., Dennis, B., & Gallagher, P. T. 2020, *ApJ*, **895**, 50

Hübner, M., Huppenkothen, D., Lasky, P. D., & Inglis, A. R. 2022, *ApJS*, **259**, 17

Huppenkothen, D., & Bachetti, M. 2022, *MNRAS*, **511**, 5689

Huppenkothen, D., Brewer, B. J., Hogg, D. W., et al. 2015, *ApJ*, **810**, 21

Huppenkothen, D., D’Angelo, C., Watts, A. L., et al. 2014c, *ApJ*, **787**, 13

Huppenkothen, D., Heil, L. M., Watts, A. L., & Göğüş, E. 2014a, *ApJ*, **795**, 11

Huppenkothen, D., Watts, A. L., & Levin, Y. 2014b, *ApJ*, **793**, 10

Huppenkothen, D., Watts, A. L., Uttley, P., et al. 2013, *ApJ*, **768**, 25

Huppenkothen, D., Younes, G., Ingram, A., et al. 2017, *ApJ*, **834**, 90

Hurley, K., Boggs, S. E., Smith, D. M., et al. 2005, *Natur*, **434**, 1098

Inglis, A. R., Ireland, J., Dennis, B. R., Hayes, L., & Gallagher, P. 2016, *ApJ*, **833**, 284

Inglis, A. R., Ireland, J., & Dominique, M. 2015, *ApJ*, **798**, 108

Inglis, A. R., Nakariakov, V. M., & Melnikov, V. F. 2008, *A&A*, **487**, 1147

Ingram, A., van der Klis, M., Middleton, M., et al. 2016, *MNRAS*, **461**, 1967

Ingram, A. R., & Motta, S. E. 2019, *NewAR*, **85**, 101524

Israel, G. L., Belloni, T., Stella, L., et al. 2005, *ApJL*, **628**, L53

Iwakiri, W., Ohno, M., Kamae, T., et al. 2010, in AIP Conf. Proc. 1279, Deciphering the Ancient Universe with Gamma-Ray Bursts (Melville, NY: AIP), 89

Jenkins, C. 2014, in AIP Conf. Ser. 1636, Bayesian Inference and Maximum Entropy Methods in Science and Engineering (Melville, NY: AIP), 106

Joachimi, B., Köhlinger, F., Handley, W., & Lemos, P. 2021, *A&A*, **647**, L5

Kane, S. R., Kai, K., Kosugi, T., et al. 1983, *ApJ*, **271**, 376

Kass, R. E., & Raftery, A. E. 1995, *J. Am. Stat. Assoc.*, **90**, 773

Kelly, B. C., Becker, A. C., Sobolewska, M., Siemiginowska, A., & Uttley, P. 2014, *ApJ*, **788**, 33

Kupriyanova, E. G., Kashapova, L. K., Reid, H. A. S., & Myagkova, I. N. 2016, *SoPh*, **291**, 3427

Levin, Y. 2006, *MNRAS*, **368**, L35

Levin, Y. 2007, *MNRAS*, **377**, 159

Levin, Y., & van Hoven, M. 2011, *MNRAS*, **418**, 659

Lindberg, C. W., Huppenkothen, D., Jones, R. L., et al. 2022, *AJ*, **163**, 29

Lucy, L. B. 2016, *A&A*, **588**, A19

Lucy, L. B. 2018, *A&A*, **614**, A25

Luger, R., Foreman-Mackey, D., & Hedges, C. 2021, *AJ*, **162**, 124

Masada, Y., Kawanaka, N., Sano, T., & Shibata, K. 2007, *ApJ*, **663**, 437

Massey, F. J. 1951, *J. Am. Stat. Assoc.*, **46**, 68

McLaughlin, J. A., Nakariakov, V. M., Dominique, M., Jelínek, P., & Takasao, S. 2018, *SSRv*, **214**, 45

Melnikov, V. F., Reznikova, V. E., Shibasaki, K., & Nakariakov, V. M. 2005, *A&A*, **439**, 727

Miller, M. C., Chirenti, C., & Strohmayer, T. E. 2019, *ApJ*, **871**, 95

Moore, C. J., Berry, C. P. L., Chua, A. J. K., & Gair, J. R. 2016, *PhRvD*, **93**, 064001

Morris, D., Battista, F., Dhuga, K., & MacLachlan, G. 2010, in AIP Conf. Proc. 1279, DECIPHERING THE ANCIENT UNIVERSE WITH GAMMA-RAY BURSTS (Melville, NY: AIP), 394

Nakariakov, V. M., Foullon, C., Myagkova, I. N., & Inglis, A. R. 2010, *ApJL*, **708**, L47

Nakariakov, V. M., & Melnikov, V. F. 2009, *SSRv*, **149**, 119

Norris, J. P., Nemiroff, R. J., Bonnell, J. T., et al. 1996, *ApJ*, **459**, 393

Palmer, D. M., Barthelmy, S., Gehrels, N., et al. 2005, *Natur*, **434**, 1107

Parks, G. K., & Winckler, J. R. 1969, *ApJL*, **155**, L117

Pastor-Marazuela, I., van Leeuwen, J., Bilous, A., et al. 2022, arXiv:2202.08002

Rasmussen, C., & Williams, C. 2006, Gaussian Processes for Machine Learning, Vol. 7 (Cambridge, NY: MIT Press)

Rea, N., Israel, G. L., Turolla, R., et al. 2009, *MNRAS*, **396**, 2419

Romero-Shaw, I. M., Talbot, C., Biscoveanu, S., et al. 2020, *MNRAS*, **499**, 3295

Samuelsson, L., & Andersson, N. 2007, *MNRAS*, **374**, 256

Savitzky, A., & Golay, M. J. E. 1964, *AnaCh*, **36**, 1627

Skilling, J. 2004, in AIP Conf. Proc. 735, 24th Int. Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, ed. R. Fischer, R. Preuss, & U. von Toussaint (Melville, NY: AIP), 395

Skilling, J. 2006, *BayAn*, **1**, 833

Smith, R. J., Ashton, G., Vajpeyi, A., & Talbot, C. 2020, *MNRAS*, **498**, 4492

Sotani, H., Iida, K., & Oyamatsu, K. 2016, *NewA*, **43**, 80

Sotani, H., Iida, K., & Oyamatsu, K. 2017, *MNRAS*, **464**, 3101

Sotani, H., Kokkotas, K. D., & Stergioulas, N. 2007, *MNRAS*, **375**, 261

Sotani, H., Kokkotas, K. D., & Stergioulas, N. 2008, *MNRAS*, **385**, L5

Speagle, J. S. 2020, *MNRAS*, **493**, 3132

Strohmayer, T. E., & Watts, A. L. 2005, *ApJL*, **632**, L111

Strohmayer, T. E., & Watts, A. L. 2006, *ApJ*, **653**, 593

Tarnopolski, M., & Marchenko, V. 2021, *ApJ*, **911**, 20

Timmer, J., & König, M. 1995, *A&A*, **300**, 707

Van Doorsselaere, T., Kupriyanova, E. G., & Yuan, D. 2016, *SoPh*, **291**, 3143

van Haasteren, R., & Vallisneri, M. 2014, *PhRvD*, **90**, 104012

VanderPlas, J. T. 2018, *ApJS*, **236**, 16

VanderPlas, J. T., & Ivezić, Ž. 2015, *ApJ*, **812**, 18

Wandelt, B. D., & Hansen, F. K. 2003, *PhRvD*, **67**, 023001

Watts, A. L. 2012, *ARA&A*, **50**, 609

Watts, A. L., & Strohmayer, T. E. 2006, *ApJL*, **637**, L117

Whittle, P. 1951, Hypothesis testing in time series analysis (Uppsala: Almqvist & Wiksell)

Wilkins, D. R. 2019, *MNRAS*, **489**, 1957

Wilson, A. G., & Nickisch, H. 2015, in 32nd Int. Conf. on Machine Learning (New York: PMLR), 1775

Zhu, X. J., & Thrane, E. 2020, *ApJ*, **900**, 14

Ziaeepour, H., & Gardner, B. 2011, *JCAP*, **12**, 001

Zimovets, I. V., McLaughlin, J. A., Srivastava, A. K., et al. 2021, *SSRv*, **217**, 66